Elucidation of respiratory microbial communities and the feasibility of breath diagnostics based on microbial volatile organic compounds (VOCs) in adult cystic fibrosis patients

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Rolf Kramer
aus Papenburg, Deutschland

1. Referent: Privatdozent Dr. Manfred G. Höfle
2. Referent: Professor Dr. Michael Steinert
eingereicht am: 24.04.2013
mündliche Prüfung (Disputation) am: 04.09.2013

Druckjahr 2013

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Kramer, R. & Abraham, W.R.: Volatile sesquiterpenes from fungi: What are they good for? Phytochemistry Reviews, 11, pp. 15-37 (2012).

Tagungsbeiträge

Kramer, R., Höfle, M.G. & Abraham, W.R.: Volatile organic compounds (VOCs) emitted by *Pseudomonas aeruginosa* and *Candida albicans*. (Poster) 3rd Joint Conference of DGHM and VAAM, Hannover (2010).

Kramer, R., Höfle, M.G. & Abraham, W.R.: Detektion von Markermolekülen (VOCs) in der Atemluft von Pneumonie-Patienten. (Poster) 25. Treffpunkt Medizintechnik, Berlin (2011).

Kramer, R., Höfle, M.G. & Abraham, W.R.: High-resolution breath analysis and volatile biomarker (VOCs) in patients with cystic fibrosis. (Oral presentation) European Meeting on Molecular Diagnostics, Scheveningen, the Hague, the Netherlands (2011).

Eines ist nur wahr neben dem anderen, und die Welt ist geräumig genug vorgesehen, um alles zu erfassen: das, was war, muß nicht von der Stelle geräumt, nur langsam verwandelt werden, so wie das, was sein wird, nicht von den Himmeln fällt im letzten Augenblick…

Rainer Maria Rilke (1875 - 1926)

...meiner Familie

Acknowledgements / Danksagung

Ein großes Dankeschön geht natürlich an meinen Mentor **Manfred Höfle** für das in mir gesetzte Vertrauen das Projekt verwirklichen zu können und für die Betreuung in den vergangenen Jahren. Ebenfalls danken für die Chance und der Überlassung dieses spannenden und herausfordernen Themas möchte ich meinem zweiten Supervisor **Wolf-Rainer Abraham**.

Ein besondere Dank gilt **Josi**, nicht nur für die tatkräftige Unterstützung im Labor sondern auch für deine liebenswerte Art und der vielen freundschaftlichen Gespräche - wäre ansonsten manchmal recht einsam gewesen im Labor… Ich möchte meinen „neuen“ und „alten“ Laborkollegen **Karsten**, **Rene**, **Marina**, **Ingrid**, **Rui**, **Douglas** und **Verena** danken - ihr alle habt mich auf eure Art unterstützt, sei es tatkräftig, mit anregenden Diskussionen oder einfach mit eurem alltäglichen Wahnsinn. Jeder von euch weiß schon wie. Ähnliches gilt auch für meine Mädels aus D0.17 **Maira**, **Andreia** und natürlich **Esther**, dir nochmals besonderen Dank für die technische Unterstützung am GC-MS. Der gesamten Arbeitsgruppe MINP möchte ich danken für die Hilfe beim NGS sequencing, insbesondere **Ruy**. Für die nette und aufgeschlossene Zusammenarbeit danke ich außerdem **Annette Sauer-Heilborn** von der MHH.

Now I have to switch the language in order to thank a number of PhDs and postdocs that enriched my life in the past years (I probably should have encouraged you earlier to learn German). **Tobi**, **Diego**, **Ignacio** - you guys are amazing, I will miss the “conferences”. Thanks to the TT-members **Giuseppe**, **Aida**, **Ivana**, **Marcia**, **Marcelo** and to all the others I first met in the corridors of the institute and who became friends: **Bahram**, **Yi-Ju**, **Marc**, **Andrew**, **Melissa** and many more. I always appreciate the international mixture and I learnt a lot from you all.

However, I want to express my special gratitude to a few close friends, I would not have succeeded without your help and support: **Alexito** and **Christian**, you both were around whenever I needed a friend; and **Daia**, you gave me drive and you made the difference in the first years - I will never forget.

Zum Schluss geht ein riesiger Dank an **meine Familie**, insbesondere **meinen Eltern**. Eure unermessliche Unterstützung gab mir immer Kraft und Sicherheit. Ihr habt mich gelehrt offen und neugierig zu sein. Es gibt keine Worte, die meine Dankbarkeit hierfür beschreiben könnten…

Summary

Cystic fibrosis (CF; mucoviscidosis) is the most common inherited disease in the western population and lung disease due to respiratory infections is the leading cause of death in CF. Routine clinical diagnostics are focusing on the detection of only a modest number of microbial pathogens but recent studies suggest that pulmonary infections in CF patients are polymicrobial. Thus, this thesis aimed to contribute to a more profound understanding of the microbial communities present in the airways of adult CF patients. For this purpose, sputum samples were collected from 56 patients and analyzed with culture-independent methods, such as single-strand conformation polymorphism (SSCP) fingerprinting and Illumina sequencing. From a smaller subcohort, samples were repeatedly collected and the temporal dynamics of microbial communities analyzed. However, spatial heterogeneity of microbial communities in the CF lung and in sputum samples are known to potentially bias the diagnostic accuracy of conventional methods based on sputum, therefore, a rapid and non-invasive breath analytical method was developed to detect microbial biomarkers (volatile organic compounds; VOCs) in exhaled air of the infected host and the feasibility of the method was tested in a CF outpatient clinic.

Only culture-independent methods revealed the broad spectrum of microorganisms associated with sputum from CF patients. Profiling of bacterial communities was performed and revealed no strong correlation between lung function and individual relative abundances of single species. In a polymicrobial perspective, however, distinct and persistent subgroups of bacterial communities could be defined by the individual compositions of species. In contrast to bacteria, an unexpected high richness of fungi was detected. Likewise, high fluctuation rates in species numbers were observed over time and between different patients, suggesting rather low colonization abilities of fungi in CF airways. For the breath analytical diagnostic approach, it was at first demonstrated that VOCs released by microorganism allowed discrimination of pathogens *in vitro*. The microbial-specific compounds observed *in vitro* were not detected in the exhaled breath of infected patients *in vivo*; however, the individual VOCs compositions monitored in exhaled breath allowed distinction between CF patients and controls.

Overall, this thesis supports the hypothesis of polymicrobial consortia being involved in pulmonary infections and provides for the first time an overview of the entire microbiome in a broader CF cohort, including fungi and bacteria. Furthermore, the potential of exhaled VOCs analysis to identify patients with certain pulmonary infections was demonstrated which may facilitate rapid and highly accurate diagnosis in the future.

Zusammenfassung

Zystische Fibrose (CF; Mukoviszidose) ist die häufigste erbliche Stoffwechselerkrankung in der europiden Bevölkerung. Die führende Todesursache bei CF ist eine durch Infektionen hervorgerufene Schädigung der Lunge. Während die klinische mikrobiologische Diagnostik noch auf einer relativ kleinen Anzahl Krankheitserreger spezialisiert ist, deuten aktuelle wissenschaftliche Studien auf polymikrobielle Lungeninfektionen als häufigste Todesursache hin. Ziel der vorliegenden Arbeit war es daher die mikrobiologischen Lebensgemeinschaften in den Atemwegen von adulten CF-Patienten zu analysieren. Sputumproben von 56 Patienten wurden mit molekularen, kulturunabhängigen Methoden untersucht, hierzu zählen molekulares "Fingerprinting" und die Sequenziertechnik von Illumina. Von einer kleineren Subkohorte wurden wiederholt Sputumproben gesammelt, um die Dynamik der mikrobiologischen Gemeinschaften zu analysieren. Bei einer auf Sputumproben basierenden Diagnostik können jedoch räumliche Heterogenitäten in den mikrobiologischen Gemeinschaften der Lunge sowie in den Proben selbst zu einem unvollständigen Befund führen. Daher wurde zusätzlich ein diagnostisches Verfahren zur Atemluftanalytik entwickelt um mikrobielle Biomarker (flüchtige organische Verbindungen; VOCs) nachzuweisen. Die Durchführbarkeit des Verfahrens wurde in der CF-Ambulanz getestet.

Die molekularen Analysen identifizierten ein breites Spektrum an Mikroorganismen im Sputum und ermöglichten die Untersuchung der genauen Zusammensetzung der unterschiedlichen bakteriellen Gemeinschaften. Eine Korrelation zwischen Lungenfunktion und der relativen Häufigkeit einzelner Arten wurde nicht festgestellt. Jedoch konnten verschiedene, teilweise persistierende Untergruppen bakterieller Gemeinschaften voneinander abgegrenzt werden, jeweils klar definiert durch ihre individuelle Zusammensetzung. Bei den Pilzen ließ sich eine unerwartet große Artenvielfalt nachweisen. Deren starke Fluktuation über die Zeit sowie zwischen verschiedenen Patienten deutet auf ein schwaches Kolonisierungspotential der Pilze hin. Für die Atemanalytik wurden verschiedener Mikroorganismen zunächst *in vitro* anhand ihrer abgesonderten VOCs unterschieden. Obwohl *in vivo* jene Verbindungen nicht detektiert wurden, ließ die individuelle Zusammensetzung der ausgeatmeten VOCs eine Differenzierung von Patienten und Kontrollen zu.

In der vorliegenden Arbeit wurden zum ersten Mal die mikrobiologischen Gemeinschaften, einschließlich Pilzen und Bakterien, in einer größeren Kohorte analysiert. Die Ergebnisse unterstützen die Hypothese von polymikrobiellen Lungeninfektionen. Das Potential der Atemanalytik zur Unterscheidung verschiedener Krankheitserreger wurde nachgewiesen und könnte die Grundlage zur Entwicklung einer verbesserten Diagnostik bilden.

Table of Contents

Chapter I
General Introduction

1.1 Microbial communities and the human microbiome

> "*Owing to this struggle for life, any variation, however slight and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, in its infinitely complex relations to other organic beings and to external nature, will tend to the preservation of that individual, and will generally be inherited by its offspring.*" (Charles Darwin, On the Origin of Species, 1859)

With these words, Charles Darwin described the major concept of nature in its entirety. The relevance of it can be seen in all aspects of biology and in particular in microbial life. In their "*struggle for life*" microorganisms conquered and adapted to the most extreme and hostile environments. Bacterial communities are accordingly found in the Antarctica at minus 13°C within the brine of isolated lakes, on the sea bottom at 350°C hot water emanated from 'black smokers' or in the toxic drainage of iron ore mining (Baross & Deming 1983) (Bond et al. 2000) (Murray et al. 2012). The evolutionary principles apply in all habitats, enable their colonization and establish the microbial diversity that can be observed in each individual environmental site. However, interactions between microorganisms are not only of competitive nature, in fact, synergistic interactions, like exchange of metabolites or mutualistic cooperation between anaerobic and aerobic bacteria, may be even more the rule than the exception (Schink 2002). Consequently, sophisticated social interrelations were elucidated within the microbial life (Velicer 2003). The complex and multifaceted outcome of these can particularly be observed in the formation of biofilms which can be regarded as integrated communities (Stoodley et al. 2002). Advanced interactions were also observed between bacteria and fungi to create selective conditions in their environment and provide growth advantages (Boer et al. 2005) (Minerdi et al. 2011). The development of defence mechanisms, invasion and colonization strategies or even antibiotic resistance are all consequences of microbial interactions including bacteria and fungi, their impact on human health cannot be overestimated (Berg et al. 2005).

Like each other environmental site, also the human body is inhabited by microorganisms. The discovery of this microbial universe in its entirety has just started. In 2008, the Human Microbiome Project was launched with the aim to elucidate the microbial diversity within healthy human beings (The Human Microbiome Project Consortium 2012a). The individual bacterial communities from different body sites, like the oral cavity and oropharynx, were presented and their specific metabolic function was demonstrated (The Human Microbiome Project Consortium 2012b). Whereas the interaction between the microbial flora and the healthy human host can be regarded as symbiosis, certain species within the flora are capable of causing damage if environmental conditions or the microbial communities are altered (Casadevall & Pirofski 2000). An overview about the outcome of host-microbe interactions are given in **Figure 1**.

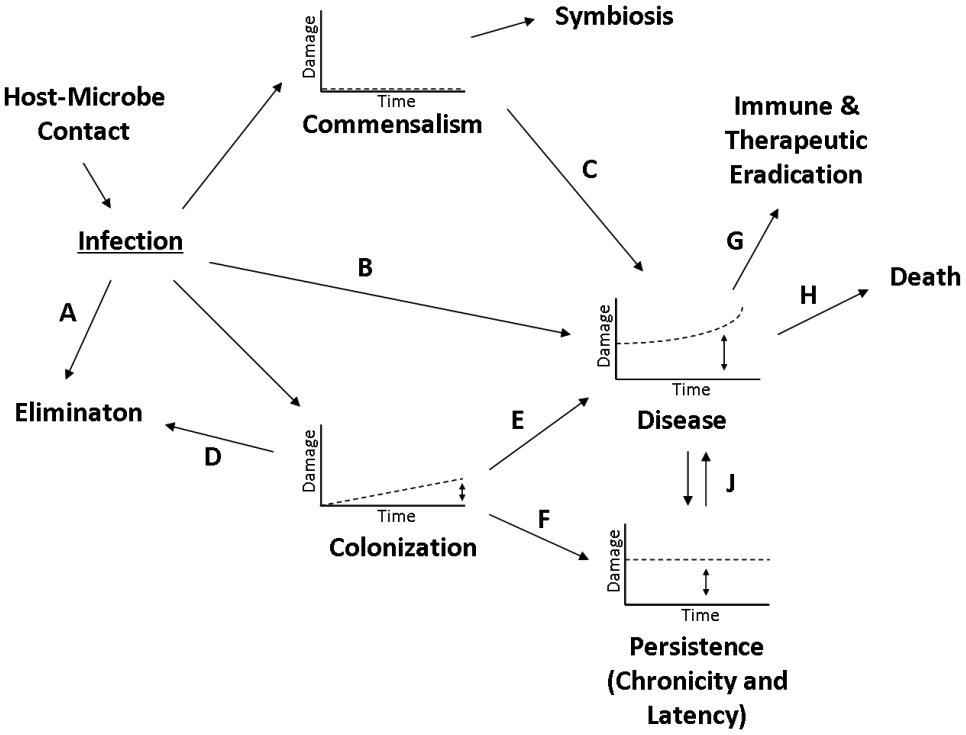


Figure 1: Outcome of infection and host-microbe interaction. In detail: A, acquisition of a microbe can be followed by elimination through physical defenses or immune mechanisms. B, acquisition results in damage and disease in the host. C, commensal microbes cause disease if the state of commensalism is disturbed by immune impairment or alterations of the host microbial flora. D, the state of colonization may be terminated by an immune response. E, the state of colonization may lead to disease if sufficient damage ensues from the interaction. F, the state of colonization may lead to a state of persistence. G, eradication of infection (but damage may be already irreversible). H, damage causes death. J, persistent infections may reactivate and cause overt disease. Double-headed arrows indicate conditions where there may be variable amounts of damage (modified from Casadevall & Pirofski 2000).

The trigger for opportunistic pathogens to cause diseases is not well understood and might be of various ecological aspects. The entire microbial community may be regarded as opportunistically pathogenic, considering the multitude of interactions between the different microbes within the flora which potentially influence each single species (Jenkinson & Lamont 2005) (Rogers et al. 2010). These "*infinitely complex relations to other organic beings and to external nature*" were defined the driver of evolution by Charles Darwin and it is becoming more and more apparent that they are similarly relevant for human health in the context of infectious diseases.

1.2 Cystic fibrosis and pathogenesis

Cystic fibrosis (CF; mucoviscidosis) is regarded as the most common genetic disorder in the caucasian population. CF affects the ion and water transport in and out of cells, causing thick mucus to build up and therefore disturb the normal function of several organs. One in 2,500 newborns has cystic fibrosis and also in other ethnic groups it is increasingly recognized (Davies et al. 2007). Albeit relatively rare, CF is of substantial public interest. Patients are treated and ask for regular check-ups in specialized centre for adult and pediatric CF health care. Depending on the individual health condition, medical services of CF patients are done ambulant and interferences in the patient's life are tempted to be kept at a minimum. Successful clinical research in the past decades led to rapid improvements in health management, therapy and diagnostics, resulting in a vast increase of life expectancy and quality for CF patients: predicted survival age, depicted in 5-year periods, revealed a median age of 28 years for 1987-1991 followed by a constant increase which is resulting in predicted survival age of 37 years for 2007-2011 (Cystic Fibrosis Foundation Patient Registry 2012). Consequently, soon more than 50 % of the population with CF will be adults, whereas in the beginning of the 1980s the number of children was still two and a half times larger than the respective number of adults (Cystic Fibrosis Foundation Patient Registry 2012). The trend is continuing and for newborns with CF in the 21st century the predicted median survival exceeds 50 years (Dodge et al. 2007). In summary, clinical manifestations of the genetic disorder and their progressions are changing with age, resulting in new challenges in healthcare of the patients. Further research is needed to meet these challenges and to retain the positive trend in survival and life quality for CF patients.

1.2.1 Genetic background

Cystic fibrosis is an autosomal recessive genetic disorder, which means both allels of a gene have to be mutated to cause the disease. Normally, it is caused by a mutation in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. The relatively high prevalence for a genetic disease might be evolutionary explained by the finding that *Salmonella typhi* uses CFTR to enter intestinal epithelial cells and the decreased susceptibility to typhoid fever might be an advantage for heterozygous carriers (Pier et al. 1998). In 1989, the cystic fibrosis gene sequence was identified on the human chromosome 7 (Rommens et al. 1989). *CFTR* encodes a protein with a size of 1,480 amino acids that serves as a channel for chloride ions and is expressed in many epithelial cell types throughout the human body (Riordan et al. 1989). Further analyses revealed that the CF mutant gene pool consists of different mutations from which the major one is responsible for over 70 % of all CF cases and defined as a nucleotide deletion resulting in the loss of one amino acid at position 508 (ΔF508) (Kerem et al. 1989). Until today, more than 1,600 different mutations duplications were described within the 189,000 base pairs of the *CFTR* locus (Quemener et al. 2010). The outcome of these mutations are multifaceted, like channel dysfunctions, abnormal production or disturbed intracellular processing of the protein (Zielenski & Tsui 1995). Correlations between the *CFTR* mutation and clinical manifestation are only known for the pancreatic functions: Carriers of homozygotes ΔF508 generally have pancreatic insufficiency (Chen & Férec 2009). The phenotypic diversity of CF is best demonstrated by the definition of the atypical or nonclassical form which is diagnosed in about 2 % of all cases: Patients demonstrate a CF like phenotype in at least one organ but have normal sweat chloride values (Boyle 2003). In most of these cases, patients have pancreatic sufficiency and lung malfunction may be less severe compared to classic CF but still exhibits phenotypic similarities. Additionally, variant CF phenotypes can also result from mutations other than *CFTR*-associated and are clinically indistinguishable from nonclassic CF (Paranjape & Zeitlin 2008).

1.2.2 Pathogenesis of lung diseases

In CF, the whole body is affected and a wide range of associated diseases in different organs are observed (Davies et al. 2007). However, respiratory infections are known to be the leading cause of death for the patients. The protein CFTR is crucial for a number of transport processes across the epithelia and may further play a role in other physiological processes (Jentsch et al. 2002). The impact of a dysfunctional protein on the airway epithelial cell

permeability is presented in **Figure 2**. Severe impairments in the ion transport lead to an increased viscosity of the mucus layer which covers human airways and serves to trap inhaled objects. In a permanent clearance process the mucus is normally removed from the respiratory tract of healthy individuals (Wanner et al. 1996). In CF patients, the disturbed ion transports in airway epithelia additionally lead to abnormal volumes of extracellular layers, resulting in a dysfunction of the complex pulmonary clearance mechanism (Matsui et al. 1998) (Boucher 2004). Together with the high viscosity of the mucus, this provides promoting conditions for microbial colonization.

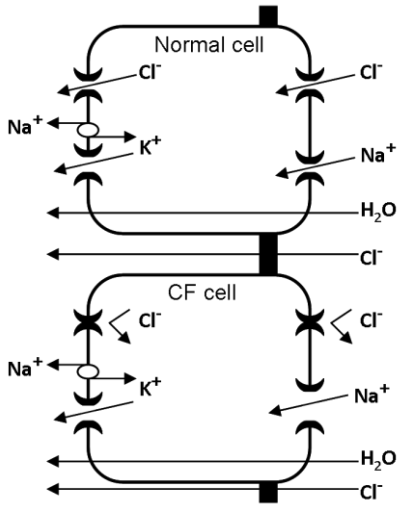


Figure 2: Comparison of models of basic electrolyte absorption properties in normal and CF airway cells. Cells are shown with tight junctions in between, basal membrane on the left and apical membrane on the right. The apical Cl^- impermeability of the CF cell depolarizes the apical membrane, but does not reverse the transmembrane electrical gradient. A significant increase in the Na^+ permeability more than compensates for the decrease in the driving force for Na^+ entry, and as Cl^- can move across this epithelium paracellularly as well as transcellularly, the net result is enhanced NaCI absorption. In turn, when fluid secretion is stimulated, Cl^- permeability of the apical membrane is not increased and normal fluid secretion characteristically fails (modified from Quinton 1990).

Since, the pulmonary airways are normally free of microbial colonization, the persistent presence of bacteria and fungi can be assumed to be wholly negative for the host (Rogers et al. 2010). Lung damage is further boosted by an increased adhesiveness and cohesiveness of mucus which is plugging small airways and lead to neutrophilic inflammation (Flume & Van Devanter 2012).

1.3 Diagnostic methods for respiratory infectious diseases in clinic and research

In a visionary article, Raoult et al. (Raoult et al. 2004) elucidated the future developments in clinical diagnostics and their impact on health management. Various culture-independent methods used in research are useful options for clinical diagnostics, like sequence-based microbial identification or mass spectrometry. Adaptation to clinical needs is crucial for this technology transfer especially the ease of operation for not highly specified users is substantial. Undoubtedly, a rapid and precise diagnosis of any infection will substantially improve the chance for patient recovery and respective individually targeted antimicrobial

therapies will diminish the costs of broad band antibiotic treatments in healthcare. This is particularly true for the complex respiratory microbial infection in CF patients. Conventional culture-based methods are still the basis of clinical diagnostics for bacteria and fungi. Although they provided the fundamentals for improvements in CF healthcare of the past decades, microbial isolation by culture is i) time-consuming, ii) selective due to species-specific media, iii) allowing at best a semi-quantitative assessment of the respective pathogen load (Rogers et al. 2003).

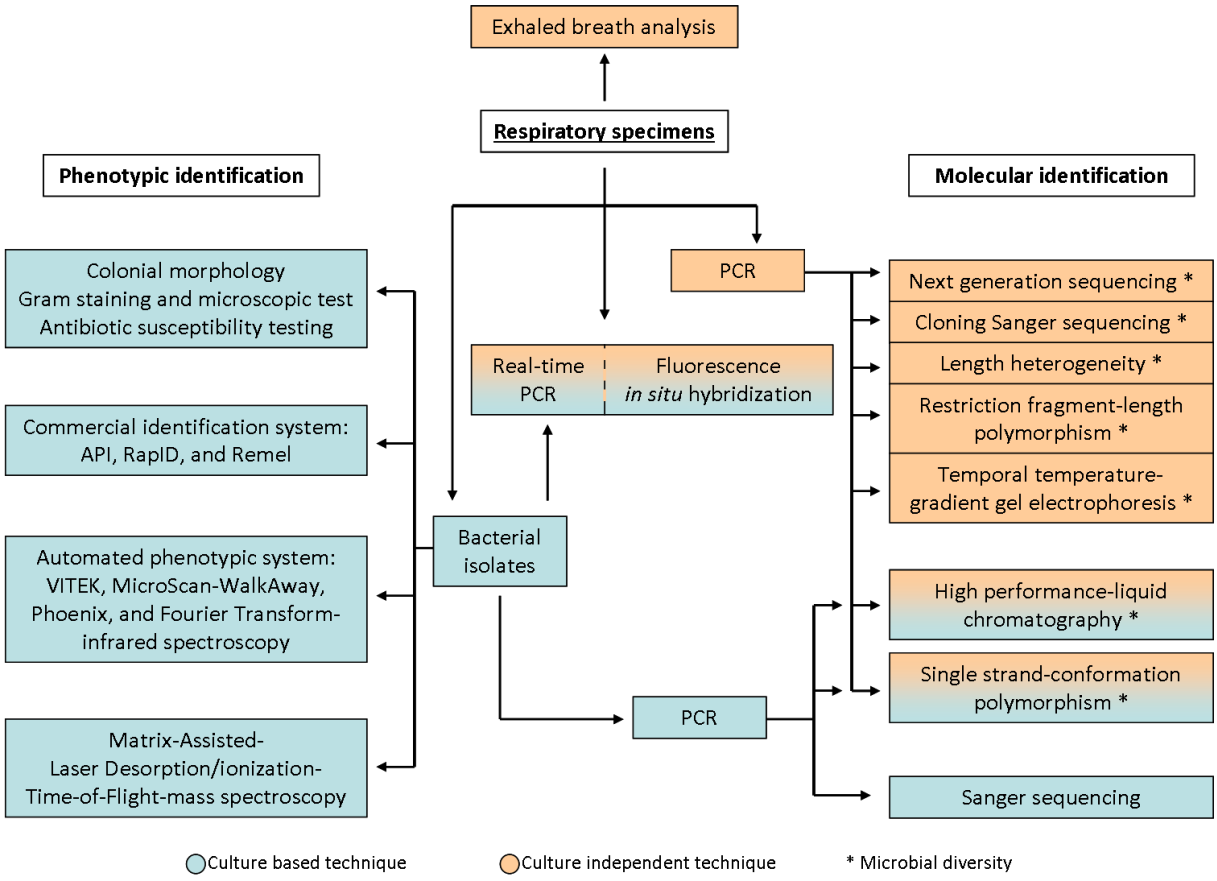


Figure 3: The different technologies used to identify bacteria in cystic fibrosis sputum samples. Conventional culture methods are useful for isolation of bacteria and they are the first step for identification and antibiotic susceptibility testing. However, especially 16S rRNA gene amplification and sequencing has revolutionized the current knowledge with the description of many emerging and multidrug-resistant bacteria that can be found in CF patients. Many other promising strategies for the identification of bacteria have been recently applied including chromatographic methods and mass spectrometry (modified from Bittar & Rolain 2010)

Technological achievements in research have further increased the knowledge about microbial infections and initiated a discussion about their polymicrobial nature (Rogers et al. 2009) (Rogers et al. 2010) (Bittar & Rolain 2010). To address these new challenges and further improve diagnostics of CF respiratory diseases, complementary technologies are needed with the potential to replace some day the conventional method of microbial isolation.

An overview about different technologies already used for accurate microbial identification in research and clinics is illustrated in **Figure 3**. The majority of the molecular identification methods are based on amplification of the rRNA genes. The bacterial 16S rRNA gene is known to have good taxonomic informativeness even in short sequence reads and enable bacterial community profiling if primers are thoroughly chosen (Soergel et al. 2012). The ribosomal 16S rRNA gene of bacterial species is known to contain highly variable regions with highly informative character for phylogenetic analysis and low variable regions to be used as target regions for universal primers to amplify 16S rRNA gene sequences in an unselective manner (Andersson et al. 2008). For fungi, the ribosomal internal transcript spacer regions (ITS) between the 18S rRNA and 28S rRNA genes were demonstrated to be best suitable for taxonomic identification (Schoch et al. 2012). Besides species-specific sequence variations, this region further exhibits high length heterogeneity across fungal species.

Major drawbacks of amplification-based methods are sequencing artefacts and bias induced by the polymerase chain reaction (PCR). High coverage rates and proof-reading enzymes minimize errors but a need for culture and amplification independent methods is apparent and would substantially improve clinical diagnostics.

1.3.1 SSCP fingerprinting

A variety of fingerprinting methods, like automated ribosomal intergenic spacer analysis (ARISA), terminal restriction fragment length polymorphism (T-RFLP), denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE) or single strand conformation polymorphism (SSCP), have been used to study microbial communities based on ribosomal DNA analysis (Nocker et al. 2007). Of these technologies, especially SSCP electrophoresis became a widely applied method in environmental, clinical and diagnostic research (Ghozzi et al. 1999) (Kumar & Shukla 2006) (Wos-Oxley et al. 2010) (Henne et al. 2012). First used to study polymorphisms and mutations in single genes, SSCP analysis was successfully applied to microbial community profiling based on the 16S rRNA gene analysis (Orita et al. 1989) (Lee et al. 1996). In this electrophoretic method, DNA-based SSCP fingerprints are achieved under non-denaturating conditions by polymorphism in the secondary/tertiary structure of single strand DNA (ssDNA). The spacial structure of nucleic acids depends on the sequence, thus DNA fragments of the same length but with different sequences fold differently and can be separated from each other (**Figure 4**). Separation of gene fragments is achieved in a native polyacrylamide gel by different migration behaviours and motilities depending on the tertiary structure. Using the 16S rRNA gene as target region,

the structure and composition of complex bacterial communities can be elucidated (Eichler et al. 2006). For fungi, the ribosomal internal transcript spacer regions (ITS) have been successfully applied to identify pathogenic fungi using this technology (Kumar & Shukla 2005).

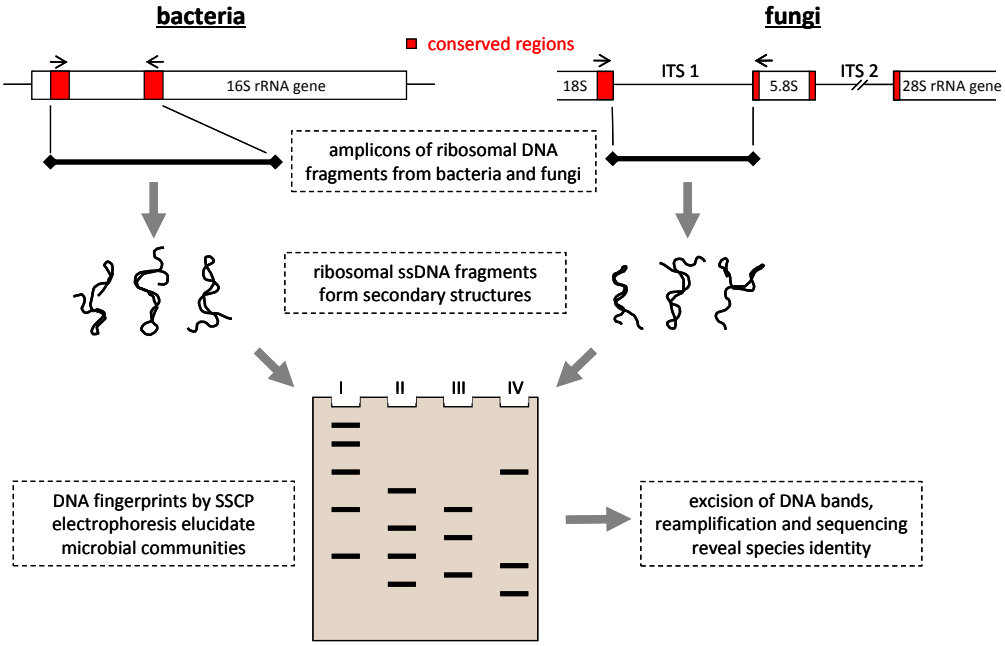


Figure 4: Schematic representation of SSCP technique for microbial community profiling. Parts of ribosomal DNA from bacteria and fungi are shown. Horizontal arrows represent exemplary primer positions for amplification of ribosomal DNA fragments with sufficient taxonomic information. Amplicons with different sequences form different secondary structures. SSCP gel with four lanes (I-IV) is indicated. All amplicons of one sample are injected in one lane, black bars represent DNA bands. Bands at the same height represent the same DNA sequence.

Though ribosomal DNA-based SSCP fingerprints allow rapid elucidation of microbial communities with minimal bioinformatic effort, the number of samples that can be analyzed and compared simultaneously is limited to 24 lanes of one gel.

1.3.2 Illumina-based next generation sequencing

Next generation sequencing (NGS) technologies have revolutionized molecular biology during the past years. Large volumes of data can be cost effectively produced within a short period of time and NGS has almost no limitation of application in any sort of sequence-based biological research. Different techniques are commercially available, like emulsion PCR-based pyrosequencing by Roche/454, solid-phase amplification-based sequencing by Illumina/Solexa or real-time single molecule sequencing by Pacific Bioscience (Bennett 2004) (Margulies et al. 2005) (Eid et al. 2009). Of these, especially the cost efficient Illumina-

based sequencing has become a widely applied technique to study bacterial diversity by 16S rRNA gene analysis (Gloor et al. 2010) (Caporaso et al. 2011) (Degnan & Ochman 2012).

Illumina sequencing, described by Metzker (Metzker 2010), is based on solid-phase amplification with cyclic reversible termination (CRT). First, DNA binds on a slide with covalently attached primers in high density which in turn defines the density of amplified template clusters. In each amplification step, the DNA polymerase incorporates only one modified nucleotides with a specific fluorescent dye before the reaction is terminated. The remaining nucleotides are washed out and fluorescent imaging is performed. In a cleavage step, the terminating group and fluorescent dye of the modified nucleotides are removed and washed out, before the cycle starts again with incorporation of the next nucleotide.

In 2008, Illumina introduced a multiplexed sequencing method in which initially prepared samples are tagged with unique short sequences (indices). Up to 12 indices, each comprised of 6 base pairs, allow identification of pooled samples which are sequenced together. The number of samples to be analyzed at once can be further increased substantially by incorporation of barcodes. These barcodes consist of unique short sequences and can be part of the PCR primers. A combinatorial approach using indices and barcodes was applied in several studies investigating microbial communities based on 16S rRNA gene analysis (Andersson et al. 2008) (Bartram et al. 2011) (Amélia Camarinha Silva 2011). Each DNA library with unique indices consists of multiple samples from different sampling sites, again each tagged with a unique barcode. Since 2,000 sequence reads per sample are sufficient to recapture the microbial community, Illumina-based sequencing is suitable for large-scale community comparisons (Caporaso et al. 2011).

1.3.3 Exhaled breath analysis based on VOC-biomarker

Routine clinical diagnostic procedures for the detection of infectious agents in the respiratory tract are based on sputum analysis. However, recent studies suggest an uneven spatial distribution for certain bacteria in sputum and a spatial heterogeneity of microbial communities in the CF lung itself (Willner et al. 2012) (Goddard et al. 2012). These regional differences may bias the diagnoses and consequently comprise a potential risk of an ineffective antibiotic treatment. In this regard, exhaled breath may be an alternative clinical specimen to be used for the detection of microbial pathogens in the lower respiratory tract. During breathing, air enters the lungs and flows through all compartments before it finally exits the lungs again. This exhaled air comprises a broad spectrum of molecules from

different origins, such as gases like oxygen, nitrogen, carbon dioxide and water but also volatile compounds produced in or outside the human body, which includes the lungs itself or other organs from where these compounds are transported via the blood stream to the alveolae (Buszewski et al. 2007) (Tisch & Haick 2010). Of particular interest for exhaled breath analysis in diagnostics of respiratory infections are volatile organic compounds (VOCs) produced and emitted by microorganisms colonizing the lower human airways. Bacteria and fungi are known to produce a broad spectrum of VOCs, like aromatic hydrocarbons, esters, alcohols or mono-, sesqui- and diterpenes (Schulz & Dickschat 2007) (Brakhage & Schroeckh 2011). Some of these VOCs have essential ecological functions for the organisms, including important roles in defence and quorum sensing (Kai et al. 2009) (O'Brien & Wright 2011) (Kramer & Abraham 2012). Therefore, these compounds may be used as specific biomarker. The diagnostic potential of these emitted VOC-biomarkers in exhaled air is attracting growing attention and different approaches were tested in recent years (Miekisch et al. 2004) (Boots et al. 2012). In the current study, a combined approach including different commercially available devices was developed in order to facilitate the crucial step of adaptation to clinical needs (**Figure 5**).

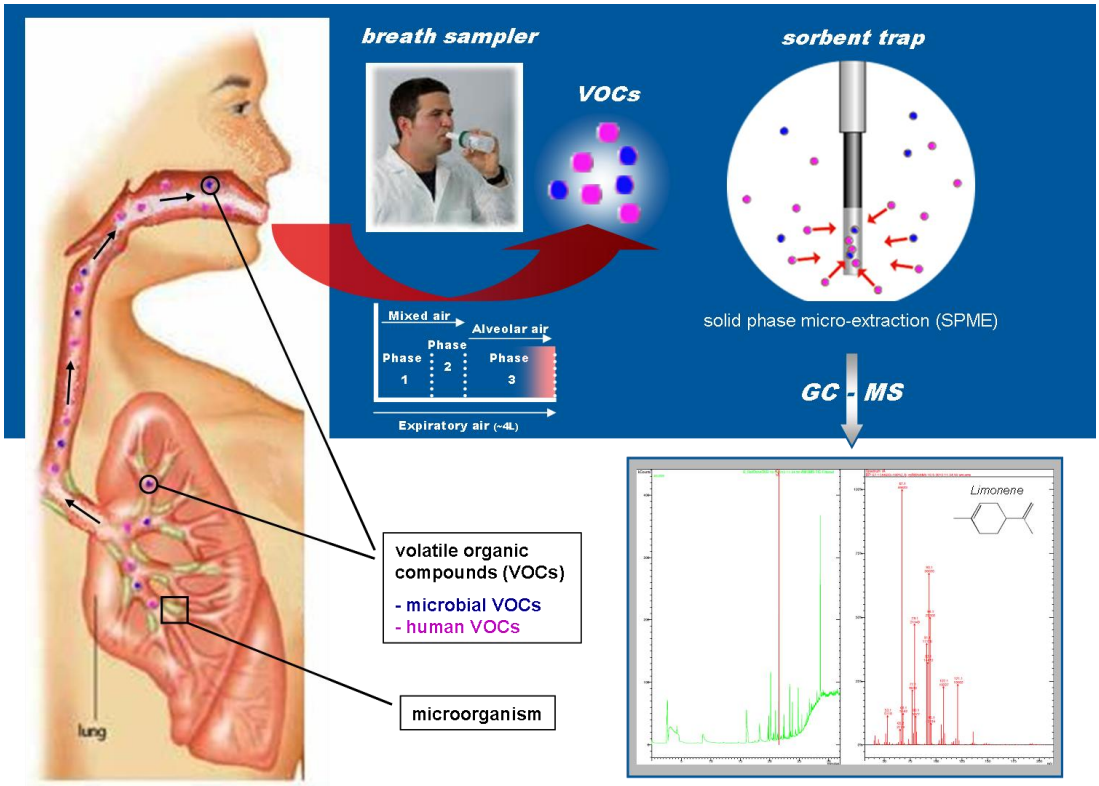


Figure 5*: Breath sampling procedure developed in the current study. VOCs are emitted by microorganisms colonizing the lower respiratory tracts. These VOCs are exhaled together with several other molecules, including

endogenous VOCs of the human body. VOCs are collected in a breath sampler, extracted by a sorbent trap (solid-phase micro-extraction) and subsequently analyzed and characterized with a GC-MS.

* Figure was designed by modifying Miekisch et al. (Miekisch et al. 2004) as well as commercial booklets of Markes International Ltd, Supelco Bellefonte and iBuySteroids.com)

By breathing through a breath sampler (BIO-VOC® sampler, Markes International Ltd, Rhonda Cynon Taff, UK), the last portion of the exhaled breath is collected. This air derived mostly from the alveolar spaces, excluding contaminants from upper airways. This sampling procedure was applied in different studies analyzing exhaled compounds (Henderson & Matthews 2002) (van den Velde et al. 2007) (Poli et al. 2010). Subsequently, a sorbent trap is exposed to the sampler and unspecifically binds the exhaled volatile organic compounds. The extraction methods is called solid phase micro-extraction (SPME), continuously developed since 1990, and applied in different breath analytical approaches as well as *in vitro* VOC emissions analysis (Arthur & Pawliszyn 1990) (Ligor et al. 2009) (Preti et al. 2009). The sorbent trap consists of a fiber coated with the polymers carboxen and polydimethylsiloxan (PDMS) and a protective needle to secure the adsorbents and the adsorbate compounds (Portable Field Sampler, Supelco Bellefonte, PA, USA). Finally, VOCs are released again by thermal desorbtion, separated and analyzed with gas chromatography coupled with mass spectrometry to characterize the different compounds.

Despite the great potential of such a fast accurate and non-invasive diagnostic method, the technical issues are demanding. The discriminative potential of VOCs as well as their origin and the influence of metabolic processes have to be elucidated. These challenges are become more apparent considering that 1,099 VOCs were recently extracted from exhaled breath of pediatric CF patients (Robroeks et al. 2010). However, different approaches are in research phase but standards in breath sampling and in VOCs extractions will have to be developed before a routine exhaled breath diagnostics for lower respiratory infections can be successfully implemented in the clinic.

1.4 Studies of microbial infections in CF respiratory tracts

In CF clinical research, microbial studies with adult patients focus mainly on the presence of single pathogens in respiratory specimen, in particular on *Pseudomonas aeruginosa* and *Staphylococcus aureus*. Certainly, *P. aeruginosa* is the most important pathogen associated with CF and several studies revealed their sophisticated adaptation to the human host (Rau et

al. 2012) (Folkesson et al. 2012). Isolated strains from CF patients exhibit remarkable phenotypic diversity, even within one individual, and evidence is growing for evolutionary progressions of single strains resulting in highly successful lineages for survival in the human environment (Mowat et al. 2011) (Folkesson et al. 2012). Interestingly, the prevalence of *P. aeruginosa* correlates with the age and is detected more often in adults, in contrast to *Haemophilus influenza* which is more prevalent in paediatric patients (Cystic Fibrosis Foundation Patient Registry 2012). Besides *P. aeruginosa*, also *S. aureus* and *Burkholderia cepacia* are extensively studied species, whereas the role of other pathogens in pulmonary infections, like *Stenotrophomonas maltophilia* or *Achromobacter xylosoxidans*, still need to be elucidated in more detail (LiPuma 2010) (Hauser et al. 2011). A number of bacteria are recognized as emerging pathogens, like species from the genera *Bordetella* and *Mycobacterium*, however, their pathogenic potential in CF remains to be shown (Olivier et al. 2003) (Spilker et al. 2008). Species from the oral cavities are commonly found in CF sputum and are suggested to immigrate into the CF lung from the upper airways which may act as a reservoir and 'stepping stone' for infectious agents (Rogers et al. 2006). Evidence for such immigrations have been particularly demonstrated for distinct clones of *P. aeruginosa* migrating from the paranasal sinuses into lower airways (Hansen et al. 2012).

With the increasing interest in anaerobic bacterial species, microbial studies shifted from single pathogen detection to microbial community profiling. With methods providing advanced resolution, a complex picture of community compositions is emerging (Bittar et al. 2008) (Guss et al. 2011). Anaerobes are observed in high numbers and their contribution to CF lung diseases is under discussion (Tunney et al. 2008) (Worlitzsch et al. 2009) (Jones 2011). As a consequence, the concept of polymicrobial consortia to be seen as one pathogenic entity was announced (Jenkinson & Lamont 2005) (Rogers et al. 2010). Such polymicrobial infections resulting in an enhanced pathogenicity are best investigated with *B. cepacia* and *P. aeruginosa*: Quorum sensing-mediated mixed biofilm formation with increased inflammatory response of the host was demonstrated (Riedel et al. 2001) (Eberl & Tümmler 2004) (Bragonzi et al. 2012). Similarly, co-infections of *P. aeruginosa* with obligate anaerobic species from the *Streptococcus milleri* group were suggested to cause severe lung diseases by enhanced virulence (Parkins et al. 2008). Determination of such polymicrobial pathogenic entities is challenging but persistence might be regarded as one key parameter for their identification. To make it more aggravating, the number of species associated with CF is increasing but long-term studies revealed relatively stable communities over time with

differences in the community structure depending on the individual (Zhao et al. 2012) (Stressmann et al. 2012). This would suggest that many species are rather persistent than transient and therefore potentially involved in the pathogenic consortia.

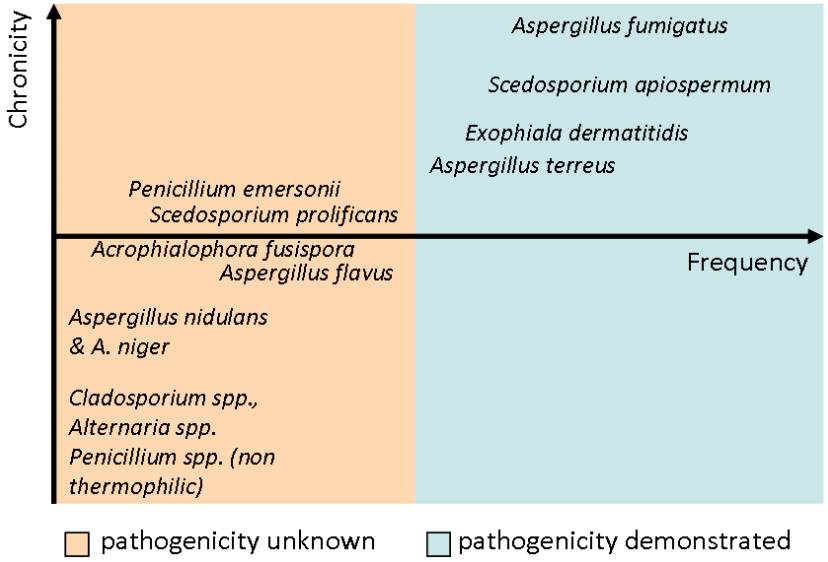


Figure 6: Filamentous fungi associated with CF according to their frequency in this clinical context and to their capacity to chronically colonize the airways (modified from Pihet et al. 2009).

Considering the whole microbial community to be one pathogenic entity also the fungal species have to be included and consequently first attempts to elucidate the complete microbiome including bacteria and fungi were made (Delhaes et al. 2012). Like bacteria, also fungal species, especially *Candida* yeasts, are commonly found in CF sputum and known to be part of the oral microbiome (Doern & Brogden-Torres 1992) (Ghannoum et al. 2010) (Müller & Seidler 2010). Likewise, they also might migrate to the lower airways. Adaptation of fungi to the CF host might be observed in *Candida dubliniensis*. This fungus has the ability to take advantage of the dehydrated respiratory secretion of CF patients and is rarely found in non-CF individuals (Hazen et al. 2001) (Peltroche-Llacsahuanga et al. 2002). The potential of *Candida* yeasts to cause primary pneumonia even in non-CF individuals is known but their incidences are low (Pasqualotto 2009). Therefore, the prevalence of *Candida* or any other fungal infections in CF patients still remains to be elucidated (Pihet et al. 2009). Few pathogenic fungal species are known, mostly filamentous fungi like *Aspergillus fumigatus*, *Scedosporium apiospermum* (teleomorph: *Pseudallescheria boydii*) and *Exophiala dermatitidis* (Müller & Seidler 2010) (LiPuma 2010). An overview about these species with known or assumed pathogenic potential is shown in **Figure 6**. Further studies are needed to analyse the fungal microbiome in CF and to elucidate the virulence of single fungal species.

1.5 Studies of exhaled breath diagnostics for lung diseases

Experimental techniques in exhaled breath analysis for clinical diagnosis and therapeutic monitoring are continuously evolving (Amann et al. 2007). The potential of exhaled breath diagnostics, is best demonstrated by the 13C-urea breath test (UBT) for the diagnoses of *Helicobacter pylori* infections, which is well established in the clinic and also considered to be sufficient to initiate treatment if results are positive (Malfertheiner et al. 2007). For lung diseases, several diagnostic approaches are currently in the research phase. In particular the detection of volatile tumor markers for the diagnosis of lung cancer is investigated. A unique chemical signature of exhaled VOCs in the breath of lung cancer patients was observed and in a recent study even dogs were trained to scent volatile markers to evaluate the diagnostic potential (Horváth et al. 2009) (Ehmann et al. 2012). Thus, a broad range of VOCs, synthesized or catabolised by cancer tissue in detectable amounts, were used in clinical trials to identify patients with different stages of lung cancer (Phillips et al. 2003) (Ligor et al. 2009) (Poli et al. 2010). Likewise, VOCs were used to distinguish between asthmatic and healthy children with high accuracy, interestingly, compounds were mainly linked with oxidative stress in these cases (Dallinga et al. 2010) (Caldeira et al. 2011). Oxidative stress and reactive oxygen species (ROS) are tightly linked with inflammation and volatile compounds generated in associated processes have the potential to be used for diagnostics (Boots et al. 2012). Likewise, the breath of CF patients was monitored for such inflammatory marker and revealed elevated levels of ethane and carbon monoxide (Paredi et al. 1999) (Paredi et al. 2000).

Despite those indirect detections of microbial infections, several studies focussed on the characterization of compounds emitted by the microorganisms itself. Great potential for microbial VOCs to be used in pathogen-directed breath testing is currently seen for *Mycobacterium tuberculosis*, *P. aeruginosa* and the fungus *A. fumigatus* (Chambers et al. 2012). A large number of patients with tuberculosis were tested for respective marker compounds and potential candidates were observed, however, technical challenges complicate a broader application (Syhre et al. 2009) (Phillips et al. 2010). *P. aeruginosa* is known for its characteristic odour, the respective compound was identified as 2-aminoacetophenone and elevated levels of it were found in the exhaled breath of infected patients (Cox & Parker 1979) (Scott-Thomas et al. 2010). Since 2-aminoacetophenone was repeatedly detected in healthy and non-infected individuals, doubts on its reliability remain. A potential marker compound was also observed for *A. fumigatus*, which causes fungal infections of the lower

respiratory tract (Chambers et al. 2011). However, the origin of that compound still has to be elucidated before a diagnostic routine can be developed.

In particular VOCs can be of various origins, being produced endogenously in the human body, they can be metabolites of food products and pharmaceuticals or inhaled from outside the body and gradually released again. Therefore, the profile of exhaled VOCs rather than single molecules might be more reliable and informative for clinical diagnostic purposes (Kharitonov & Barnes 2001).

1.6 Statistical analysis

New generations of bio- and chemo-analytical technologies are characterized by low detection limits and precise data recording. Large data sets are produced and require advanced mathematical tools, like multivariant analyses, to be explored thoroughly. Bio-statistical procedures are becoming more important and are often essential to find patterns between samples or variables in large data sets or to draw the right conclusions. In the current studies, technologies like Illumina-based sequencing for microbial community profiling and SPME/GC-MS for analytical screening of exhaled breath were applied to clinical samples. The resulting data sets were analyzed and interpreted by the use of PRIMER 6 (Version.6.1.6, PRIMER-E, Plymouth Marine Laboratory, UK). Important statistical routines being used are introduced below and additionally described in the respective chapters.

• Multi-dimensional scaling (MDS) plots: Multivariante statistical techniques use clustering or ordination to explore patterns within the data sets. In the current studies, MDS plots were mostly used to display the similarity between a set of samples. Similarity or distances between every pair of samples were calculated by the use of coefficients, like Euclidean distance or Bray-Curtis similarity. The resultant resemblance or (rank) similarity matrices express the respective scores for each pair (K. R. Clarke & M. Ainsworth 1993). The ideal display of the similarities or distances according to these scores would be in *n*-dimensional space, but was reduced in two dimensional space for an intuitively accessible manner. The 2D stress value, given for every MDS plot, basically measures these differences between ideal distances in higher dimensional space and the actual distances in two dimensional space. Therefore, the more similar the samples are according to the pairwise similarity calculations, the more close together the samples are displayed in the MDS plot. Single variables can be superimposed on

the plots for demonstration purposes, however, for the construction of the MDS plots all variables are always taken into account.

• RELATE: RELATE is a comparative (Mantel-type) test, which measures the similarity between two resemblance matrices (Clarke & Warwick 2001). In the current study, similarity was measured by the Spearman's rank correlation coefficient Rho and data sets were permuted to test for null hypothesis. Histograms were displayed to show distribution of Rho values for permutations as well as Rho value for the actual data sets.

• Similarity profile routine (SIMPROF): SIMPROF is analysing an *a priori* unstructured data set in order to test for the presence of sample groups (Clarke et al. 2008). In the current study, it was used to interpret a dendrogram from hierarchical cluster analysis of a resemblance matrix. SIMPROF is a permutation procedure and avoids over-interpretation of cluster analysis.

1.7 Research approach

Cystic fibrosis is becoming a more complex disease in adults. Recent developments in health care increased the life expectancy dramatically and new challenges for clinical research have to be addressed in order to facilitate further improvements in diagnostics as well as treatment for CF patients. Ecological studies of microbial communities and respective correlation analysis with clinical factors are crucial for the understanding of adult CF lung diseases and similarly crucial for future developments in diagnostics.

On the one hand, the pathogenic potential of emerging species has to be elucidated which includes the identification of polymicrobial pathogenic consortia as well as elucidation of the general prevalence of fungi in the respiratory tract. On the other hand, healthcare of CF patients will further benefit from advanced diagnostic methods suitable for clinical use that avoid bias due to sputum preparation. Both challenges were addressed in the current study (**Figure 7**). Bacterial and fungal communities were elucidated with culture-independent methods and analyzed with regard to host factors as well as clinical features of the patients. Additionally, a breath analytical diagnostic approach was developed and its feasibility for clinical use investigated in order to detect volatile microbial biomarker *in vivo* in the exhaled breath of CF patients.

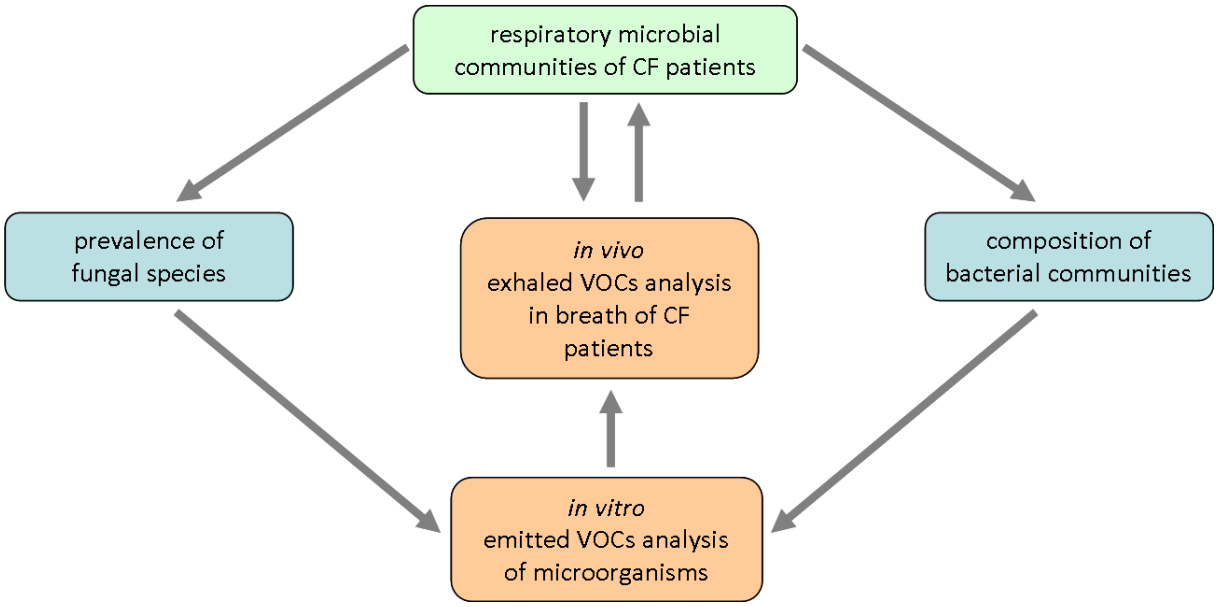


Figure 7: Project scheme of the current study. Diagnostics of respiratory communities associated with cystic fibrosis was addressed in two ways: Elucidation of fungal prevalence as well as bacterial community compositions in sputum samples and monitoring of microbial VOCs in exhaled breath of CF patients. Microbial VOCs of prominent microorganism in CF sputum were first characterized *in vitro*, including bacteria and fungi. Detection of microbial VOCs in exhaled breath entails potential for a non-invasive and rapid diagnosis of respiratory infections.

1.7.1 Objectives of thesis

Three objectives were defined in order to address the main goal of the current thesis, which is to further improve the technology and comprehension in diagnostics of lower respiratory tract infections on the basis of a more profound understanding of the microbial communities in the airways of CF patients. Therefore, the following aims were approached:

- Chapter II – To elucidate the structure and composition of bacterial communities in sputum samples, including the validation of different diagnostic methods for microbial profiling and assessment of the individual pathogenic potential of certain bacterial species for CF patients.

- Chapter III – To identify and analyse the fungal microbiome present in sputum samples, consequentially compare with bacterial taxa and estimate the prevalence and role of fungal pathogens in respiratory tract infections of CF patients.

- Chapter IV – To develop a rapid, non-invasive diagnostic method based on the detection of exhaled VOCs in the human breath, determine VOC-biomarker emitted by pulmonary pathogens *in-vitro* and test the feasibility of the method in the CF outpatient clinic *in-vivo*.

1.8 References

Amann A, Spanel P & Smith D (2007) Breath Analysis: The Approach Towards Clinical Applications. *Mini Reviews in Medicinal Chemistry* 7: 115–129.

Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P & Engstrand L (2008) Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing. *PLoS ONE* 3: e2836.

Arthur CL & Pawliszyn J (1990) Solid phase microextraction with thermal desorption using fused silica optical fibers. *Anal. Chem.* 62: 2145–2148.

Baross JA & Deming JW (1983) Growth of "black smoker" bacteria at temperatures of at least 250 °C. *Nature (London)* 303: 423–426.

Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G & Neufeld JD (2011) Generation of multi-million 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl. Environ. Microbiol.* 77: 3846–3852.

Bennett S (2004) Solexa Ltd. *Pharmacogenomics* 5: 433–438.

Berg G, Eberl L & Hartmann A (2005) The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ. Microbiol.* 7: 1673–1685.

Bittar F, Richet H, Dubus J-C, Reynaud-Gaubert M, Stremler N, Sarles J, Raoult D & Rolain J-M (2008) Molecular Detection of Multiple Emerging Pathogens in Sputa from Cystic Fibrosis Patients. *PLoS ONE* 3: e2908

Bittar F & Rolain J-M (2010) Detection and accurate identification of new or emerging bacteria in cystic fibrosis patients. *Clin. Microbiol. Infect.* 16: 809–820.

Boer W de, Folman LB, Summerbell RC & Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol. Rev.* 29: 795–811.

Bond PL, Smriga SP & Banfield JF (2000) Phylogeny of Microorganisms Populating a Thick, Subaerial, Predominantly Lithotrophic Biofilm at an Extreme Acid Mine Drainage Site. *Appl Environ Microbiol* 66: 3842–3849.

Boots AW, van Berkel JJBN, Dallinga JW, Smolinska A, Wouters EF & van Schooten FJ (2012) The versatile use of exhaled volatile organic compounds in human health and disease. *J Breath Res* 6: 027108.

Boucher RC (2004) New concepts of the pathogenesis of cystic fibrosis lung disease. *Eur. Respir. J.* 23: 146–158.

Boyle MP (2003) Nonclassic cystic fibrosis and CFTR-related diseases. *Curr Opin Pulm Med* 9: 498–503.

Bragonzi A et al. (2012) Modelling co-infection of the cystic fibrosis lung by Pseudomonas aeruginosa and Burkholderia cenocepacia reveals influences on biofilm formation and host response. *PLoS ONE* 7: e52330.

Brakhage AA & Schroeckh V (2011) Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genet. Biol.* 48: 15–22.

Buszewski B, Kęsy M, Ligor T & Amann A (2007) Human exhaled air analytics: biomarkers of diseases. *Biomedical Chromatography* 21: 553–566.

Caldeira M, Barros AS, Bilelo MJ, Parada A, Câmara JS & Rocha SM (2011) Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J Chromatogr A* 1218: 3771–3780.

Amélia Camarinha Silva (2011) *Dissertation:* Bacterial community structure and temporal dynamics of the human anterior nares - Chapter IV. Technischen Universität Carolo‐Wilhelmina, Braunschweig.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N & Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108 Suppl 1: 4516–4522.

Casadevall A & Pirofski L-A (2000) Host-Pathogen Interactions: Basic Concepts of Microbial Commensalism, Colonization, Infection, and Disease. *Infect. Immun.* 68: 6511–6518.

Chambers ST, Bhandari S, Scott-Thomas A & Syhre M (2011) Novel diagnostics: progress toward a breath test for invasive Aspergillus fumigatus. *Med. Mycol.* 49 Suppl 1: S54–61.

Chambers ST, Scott-Thomas A & Epton M (2012) Developments in novel breath tests for bacterial and fungal pulmonary infection. *Curr Opin Pulm Med* 18: 228–232.

Chen J-M & Férec C (2009) Chronic pancreatitis: genetics and pathogenesis. *Annu Rev Genomics Hum Genet* 10: 63–87.

Clarke KR, Somerfield PJ & Gorley RN (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology* 366: 56–69.

Clarke KR & Warwick RM (2001) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*. PRIMER-E Limited.

Cox CD & Parker J (1979) Use of 2-aminoacetophenone production in identification of Pseudomonas aeruginosa. *J. Clin. Microbiol.* 9: 479–484.

Cystic Fibrosis Foundation Patient Registry (2012) *2011 Annual Data Report*. Cystic Fibrosis Foundation, Bethesda, Maryland http://www.cff.org/UploadedFiles/research/ClinicalResearch/2011-Patient-Registry.pdf (Accessed January 8, 2013).

Dallinga JW, Robroeks CMHHT, Van Berkel JJBN, Moonen EJC, Godschalk RWL, Jöbsis Q, Dompeling E, Wouters EFM & Van Schooten FJ (2010) Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clinical & Experimental Allergy* 40: 68–76.

Davies JC, Alton EWFW & Bush A (2007) Cystic fibrosis. *BMJ* 335: 1255–1259.

Degnan PH & Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J* 6: 183–194.

Delhaes L et al. (2012) The Airway Microbiota in Cystic Fibrosis: A Complex Fungal and Bacterial Community—Implications for Therapeutic Management. *PLoS ONE* 7: e36313.

Dodge JA, Lewis PA, Stanton M & Wilsher J (2007) Cystic fibrosis mortality and survival in the UK: 1947-2003. *Eur. Respir. J.* 29: 522–526.

Doern GV & Brogden-Torres B (1992) Optimum use of selective plated media in primary processing of respiratory tract specimens from patients with cystic fibrosis. *J Clin Microbiol* 30: 2740–2742.

Eberl L & Tümmler B (2004) Pseudomonas aeruginosa and Burkholderia cepacia in cystic fibrosis: genome evolution, interactions and adaptation. *Int. J. Med. Microbiol.* 294: 123–131.

Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G & Walles T (2012) Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. *Eur. Respir. J.* 39: 669–676.

Eichler S, Christen R, Höltje C, Westphal P, Bötel J, Brettar I, Mehling A & Höfle MG (2006) Composition and Dynamics of Bacterial Communities of a Drinking Water Supply System as Assessed by RNA- and DNA-Based 16S rRNA Gene Fingerprinting. *Appl Environ Microbiol* 72: 1858–1872.

Eid J et al. (2009) Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323: 133–138.

Flume PA & Van Devanter DR (2012) State of progress in treating cystic fibrosis respiratory disease. *BMC Med* 10: 88.

Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N & Molin S (2012) Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. *Nat. Rev. Microbiol.* 10: 841–851.

Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A & Gillevet PM (2010) Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathog* 6: e1000713.

Ghozzi R et al. (1999) Capillary Electrophoresis–Single-Strand Conformation Polymorphism Analysis for Rapid Identification of Pseudomonas aeruginosa and Other Gram-Negative Nonfermenting Bacilli Recovered from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 37: 3374–3379.

Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R & Reid G (2010) Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS ONE* 5: e15406.

Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL & Singh PK (2012) Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 109: 13769–13774.

Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S & Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 5: 20–29.

Hansen SK et al. (2012) Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *ISME J* 6: 31–45.

Hauser AR, Jain M, Bar-Meir M & McColley SA (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clin Microbiol Rev* 24: 29–70.

Hazen KC, Wu JG & Masuoka J (2001) Comparison of the Hydrophobic Properties of Candida albicans and Candida dubliniensis. *Infect Immun* 69: 779–786.

Henderson KA & Matthews IP (2002) Biological monitoring of midwives' exposure to N2O using the Bio-VOC breath sampler. *Journal of Exposure Science and Environmental Epidemiology* 12: 309–312.

Henne K, Kahlisch L, Brettar I & Höfle MG (2012) Analysis of Structure and Composition of Bacterial Core Communities in Mature Drinking Water Biofilms and Bulk Water of a Citywide Network in Germany. *Appl. Environ. Microbiol.* 78: 3530–3538.

Horváth I, Lázár Z, Gyulai N, Kollai M & Losonczy G (2009) Exhaled biomarkers in lung cancer. *Eur Respir J* 34: 261–275.

Jenkinson HF & Lamont RJ (2005) Oral microbial communities in sickness and in health. *Trends Microbiol.* 13: 589–595.

Jentsch TJ, Stein V, Weinreich F & Zdebik AA (2002) Molecular structure and physiological function of chloride channels. *Physiol. Rev.* 82: 503–568.

Jones AM (2011) Anaerobic bacteria in cystic fibrosis: pathogens or harmless commensals? *Thorax* 66: 558–559.

K. R. Clarke & M. Ainsworth (1993) A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 205–219.

Kai M, Haustein M, Molina F, Petri A, Scholz B & Piechulla B (2009) Bacterial volatiles and their action potential. *Applied Microbiology and Biotechnology* 81: 1001–1012.

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M & Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* 245: 1073–1080.

Kharitonov SA & Barnes PJ (2001) Exhaled markers of inflammation. *Curr Opin Allergy Clin Immunol* 1: 217–224.

Kramer R & Abraham W-R (2012) Volatile sesquiterpenes from fungi: what are they good for? *Phytochemistry Reviews* 11: 15–37.

Kumar M & Shukla PK (2006) Single-stranded conformation polymorphism of large subunit of ribosomal RNA is best suited to diagnosing fungal infections and differentiating fungi at species level. *Diagnostic Microbiology and Infectious Disease* 56: 45–51.

Kumar M & Shukla PK (2005) Use of PCR targeting of internal transcribed spacer regions and single-stranded conformation polymorphism analysis of sequence variation in different regions of rrna genes in fungi for rapid diagnosis of mycotic keratitis. *J. Clin. Microbiol.* 43: 662–668.

Lee DH, Zo YG & Kim SJ (1996) Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. *Appl. Environ. Microbiol.* 62: 3112–3120.

Ligor M et al. (2009) Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. *Clin. Chem. Lab. Med.* 47: 550–560.

LiPuma JJ (2010) The Changing Microbial Epidemiology in Cystic Fibrosis. *Clin. Microbiol. Rev.* 23: 299–323.

Malfertheiner P et al. (2007) Current concepts in the management of Helicobacter pylori infection: the Maastricht III Consensus Report. *Gut* 56: 772–781.

Margulies M et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW & Boucher RC (1998) Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 95: 1005–1015.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11: 31–46.

Miekisch W, Schubert JK & Noeldge-Schomburg GF. (2004) Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clinica Chimica Acta* 347: 25–39.

Minerdi D, Bossi S, Maffei ME, Gullino ML & Garibaldi A (2011) Fusarium oxysporum and its bacterial consortium promote lettuce growth and expansin A5 gene expression through microbial volatile organic compound (MVOC) emission. *FEMS Microbiology Ecology* 76: 342–351.

Mowat E, Paterson S, Fothergill JL, Wright EA, Ledson MJ, Walshaw MJ, Brockhurst MA & Winstanley C (2011) Pseudomonas aeruginosa Population Diversity and Turnover in Cystic Fibrosis Chronic Infections. *Am. J. Respir. Crit. Care Med.* 183: 1674–1679.

Müller F-MC & Seidler M (2010) Characteristics of pathogenic fungi and antifungal therapy in cystic fibrosis. *Expert Rev Anti Infect Ther* 8: 957–964.

Murray AE et al. (2012) Microbial life at −13 °C in the brine of an ice-sealed Antarctic lake. *PNAS* USA 109:20626–20631.

Nocker A, Burr M & Camper AK (2007) Genotypic Microbial Community Profiling: A Critical Technical Review. *Microb Ecol* 54: 276–289.

O'Brien J & Wright GD (2011) An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* 22: 552–558.

Olivier KN et al. (2003) Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* 167: 828–834.

Orita M, Iwahana H, Kanazawa H, Hayashi K & Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *PNAS* 86: 2766–2770.

Paranjape SM & Zeitlin PL (2008) Atypical cystic fibrosis and CFTR-related diseases. *Clin Rev Allergy Immunol* 35: 116–123.

Paredi P, Kharitonov SA, Leak D, Shah PL, Cramer D, Hodson ME & Barnes PJ (2000) Exhaled ethane is elevated in cystic fibrosis and correlates with carbon monoxide levels and airway obstruction. *Am. J. Respir. Crit. Care Med.* 161: 1247–1251.

Paredi P, Shah PL, Montuschi P, Sullivan P, Hodson ME, Kharitonov SA & Barnes PJ (1999) Increased carbon monoxide in exhaled air of patients with cystic fibrosis. *Thorax* 54: 917–920.

Parkins MD, Sibley CD, Surette MG & Rabin HR (2008) The Streptococcus milleri group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatr. Pulmonol.* 43: 490–497.

Pasqualotto AC (2009) Candida and the paediatric lung. *Paediatr Respir Rev* 10: 186–191.

Peltroche-Llacsahuanga H, Döhmen H & Haase G (2002) Recovery of Candida dubliniensis from sputum of cystic fibrosis patients. *Mycoses* 45: 15–18.

Phillips M, Basa-Dalay V, Bothamley G, Cataneo RN, Lam PK, Natividad MPR, Schmitt P & Wai J (2010) Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb)* 90: 145–151.

Phillips M, Cataneo RN, Cummin ARC, Gagliardi AJ, Gleeson K, Greenberg J, Maxfield RA & Rom WN (2003) Detection of Lung Cancer With Volatile Markers in the Breath*. *CHEST* 123: 2115–2123.

Pier GB, Grout M, Zaidi T, Meluleni G, Mueschenborn SS, Banting G, Ratcliff R, Evans MJ & Colledge WH (1998) Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature* 393: 79–82.

Pihet M et al. (2009) Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis – a review. *Medical Mycology* 47: 387–397.

Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, Casalini A & Mutti A (2010) Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 878: 2643–2651.

Preti G, Thaler E, Hanson CW, Troy M, Eades J & Gelperin A (2009) Volatile compounds characteristic of sinus-related bacteria and infected sinus mucus: analysis by solid-phase microextraction and gas chromatography-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 877: 2011–2018.

Quemener S et al. (2010) Complete ascertainment of intragenic copy number mutations (CNMs) in the CFTR gene and its implications for CNM formation at other autosomal loci. *Hum. Mutat.* 31: 421–428.

Quinton PM (1990) Cystic fibrosis: a disease in electrolyte transport. *FASEB J.* 4: 2709–2717.

Raoult D, Fournier PE & Drancourt M (2004) What does the future hold for clinical microbiology? *Nat. Rev. Microbiol.* 2: 151–159.

Rau MH, Marvig RL, Ehrlich GD, Molin S & Jelsbak L (2012) Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. *Environ. Microbiol.* 14: 2200–2211.

Riedel K et al. (2001) N-acylhomoserine-lactone-mediated communication between Pseudomonas aeruginosa and Burkholderia cepacia in mixed biofilms. *Microbiology (Reading, Engl.)* 147: 3249–3262.

Riordan JR et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245: 1066–1073.

Robroeks CMHHT et al. (2010) Metabolomics of Volatile Organic Compounds in Cystic Fibrosis Patients and Controls. *Pediatric Research* 68: 75–80.

Rogers GB, Carroll MP & Bruce KD (2009) Studying bacterial infections through culture-independent approaches. *J Med Microbiol* 58: 1401–1418.

Rogers GB, Carroll MP, Hoffman LR, Walker AW, Fine DA & Bruce KD (2010) Comparing the microbiota of the cystic fibrosis lung and human gut. *Gut Microbes* 1: 85–93.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ & Bruce KD (2006) Use of 16S rRNA Gene Profiling by Terminal Restriction Fragment Length Polymorphism Analysis To Compare Bacterial Communities in Sputum and Mouthwash Samples from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 44: 2601–2604.

Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ & Bruce KD (2003) Bacterial Diversity in Cases of Lung Infection in Cystic Fibrosis Patients: 16S Ribosomal DNA (rDNA) Length Heterogeneity PCR and 16S rDNA Terminal Restriction Fragment Length Polymorphism Profiling. *J. Clin. Microbiol.* 41: 3548–3558.

Rommens JM et al. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 245: 1059–1065.

Schink B (2002) Synergistic interactions in the microbial world. *Antonie Van Leeuwenhoek* 81: 257–261.

Schoch CL et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* doi: 10.1073/ pnas.1117018109..

Schulz S & Dickschat JS (2007) Bacterial volatiles: the smell of small organisms. *Nat Prod Rep* 24: 814–842.

Scott-Thomas A, Syhre M, Pattemore P, Epton M, Laing R, Pearson J & Chambers S (2010) 2-Aminoacetophenone as a potential breath biomarker for Pseudomonas aeruginosa in the cystic fibrosis lung. *BMC Pulmonary Medicine* 10: 56.

Soergel DAW, Dey N, Knight R & Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* 6: 1440–1444.

Spilker T, Liwienski AA & LiPuma JJ (2008) Identification of Bordetella spp. in respiratory specimens from individuals with cystic fibrosis. *Clin. Microbiol. Infect.* 14: 504–506.

Stoodley P, Sauer K, Davies DG & Costerton JW (2002) Biofilms as Complex Differentiated Communities. *Annual Review of Microbiology* 56: 187–209.

Stressmann FA et al. (2012) Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* 67: 867–873.

Syhre M, Manning L, Phuanukoonnon S, Harino P & Chambers ST (2009) The scent of Mycobacterium tuberculosis--part II breath. *Tuberculosis (Edinb)* 89: 263–266.

The Human Microbiome Project Consortium (2012a) A framework for human microbiome research. *Nature* 486: 215–221.

The Human Microbiome Project Consortium (2012b) Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207–214.

Tisch U & Haick H (2010) Arrays of chemisensitive monolayer-capped metallic nanoparticles for diagnostic breath testing. *Reviews in Chemical Engineering* 26: 171–179.

Tunney MM et al. (2008) Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* 177: 995–1001.

Van den Velde S, Quirynen M, Van hee Paul & van Steenberghe D (2007) Halitosis associated volatiles in breath of healthy subjects. *Journal of Chromatography B* 853: 54–61.

Velicer GJ (2003) Social strife in the microbial world. *Trends Microbiol.* 11: 330–337.

Wanner A, Salathé M & O'Riordan TG (1996) Mucociliary clearance in the airways. *Am. J. Respir. Crit. Care Med.* 154: 1868–1902.

Willner D, Haynes MR, Furlan M, Schmieder R, Lim YW, Rainey PB, Rohwer F & Conrad D (2012) Spatial distribution of microbial communities in the cystic fibrosis lung. *ISME J* 6: 471–474.

Worlitzsch D, Rintelen C, Böhm K, Wollschläger B, Merkel N, Borneff-Lipp M & Döring G (2009) Antibiotic-resistant obligate anaerobes during exacerbations of cystic fibrosis patients. *Clin. Microbiol. Infect.* 15: 454–460.

Wos-Oxley ML, Plumeier I, Eiff C von, Taudien S, Platzer M, Vilchez-Vargas R, Becker K & Pieper DH (2010) A poke into the diversity and associations within human anterior nare microbial communities. *The ISME Journal* 4: 839–851.

Zhao J et al. (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci. U.S.A.* 109: 5809–5814.

Zielenski J & Tsui LC (1995) Cystic fibrosis: genotypic and phenotypic variations. *Annu. Rev. Genet.* 29: 777–807.

Chapter II

Assessment of distinct bacterial communities and their pathogenic potential in sputum from adult cystic fibrosis patients

Rolf Kramer[1], Annette Sauer-Heilborn[2], Ruy Jáuregui1[3],
Carlos A. Guzman[1], Manfred G. Höfle[1]

[1]Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2]Department of Pneumology, Hannover Medical School (MHH), Germany

[3]Department of Medical Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany

2.1 Abstract

Lung disease due to microbial infections is the leading cause of death in cystic fibrosis (CF) patients. Although routine clinical diagnostics is focusing on only a modest number of species, recent studies suggest that infections are polymicrobial. In the current study, sputum samples of 56 adult CF patients were analyzed and different diagnostic methods compared, including cultivation, 16S rRNA gene analysis based on Illumina sequencing and single-strand conformation polymorphism (SSCP) fingerprinting. Occurrence frequencies of known pathogens observed with cultivation and culture-independent methods were highly similar; however, only molecular approaches revealed the broad spectrum of bacteria associated with CF. Furthermore, no strong correlation was observed between relative abundances of single bacterial species and lung functions of the infected patients. A polymicrobial perspective was utilized and multivariant statistical analysis revealed distinct subgroups of bacterial communities in the cohort defined by their individual compositions. Persistence of distinct communities was then demonstrated by a longitudinal analysis of 13 CF patients. Characterization of such polymicrobial consortia involved in respiratory infections may further improve diagnostics of lung diseases and health care for CF patients.

2.2 Introduction

Lower respiratory infections are among the five most common causes of death worldwide and are the leading cause for patients with cystic fibrosis. In the Caucasian population, one out of 2,500 newborns has cystic fibrosis (CF) and in other ethnic groups its importance is increasing, what makes the disease a substantial public health issue (Davies et al. 2007). CF is an autosomal recessive genetic disorder, caused by mutations in the gene for a chloride ion channel (CFTR) which result in a dysfunction of the mucociliary clearance mechanisms and in conditions that promote microbial growth in the human lungs (Matsui et al. 1998) (Boucher 2004). In particular bacteria cause lower respiratory infections in CF with recurrent exacerbations and entail immune responses which in turn are thought to be responsible for the majority of irreversible lung damages (Sibley et al. 2008).

On the one hand for CF patients improved diagnostics and health management have lead to a constantly increasing life expectancy and quality of patients, on the other hand it is becoming apparent that disease manifestations and infections are changing with age (Davies et al. 2007) (Cystic Fibrosis Foundation Patient Registry 2012). These new challenges have to be addressed and alternatives to conventional microbial culture-based diagnostics were developed, including molecular fingerprinting methods and, more recently, next generation sequencing technologies (Ghozzi et al. 1999) (Bittar & Rolain 2010). Hereby, bacterial communities are elucidated via 16S rRNA gene analysis and revealed a broad spectrum of microbial species associated with CF, many of which were previously under-diagnosed or even not undetected (Rogers et al. 2004) (Guss et al. 2011). A complex picture of bacterial communities in the CF respiratory tracts is developing with emerging pathogens, anaerobic species and polymicrobial interactions (Bittar et al. 2008) (Tunney et al. 2008) (Sibley et al. 2008) (Worlitzsch et al. 2009). However, their potential to cause lower respiratory infections in CF remains to be unclear for most species. In this regard, multiple factor analyses were performed with distinct environmental or physiological parameter associated with the host and the airway microbiota in order to reveal any relation (Cox et al. 2010) (Klepac-Ceraj et al. 2010). As a consequence, improved understanding of the complex bacterial community rather than the identification of single pathogenic species may positively influence treatment and health management of patients (Flanagan et al. 2007). In this context, Rogers et al. (2009) suggested to consider the bacterial community in CF airways as one pathogenic entity where the virulence of single species may be triggered by co-colonization and interactions with other bacteria in a polymicrobial consortium. Persistence of bacteria over time may give insights in

their contribution to the respiratory communities that infect the CF airways. Recent respective investigations revealed stable communities over time within patients including species not classically considered to be pathogenic in CF (Stressmann et al. 2012). The same study observed a higher diversity among patients rather than within the same individual which demonstrates that each CF airway infection has a unique character and treatment strategies have to be individually adjusted.

However, introduction of molecular diagnostic methods like pyrosequencing in routine clinical diagnostics remain to be challenging. Until then, elucidation of emerging pathogens and associated polymicrobial consortia by molecular methods will facilitate clinical attention for them and facilitate their incorporation in routine diagnostics. In the current study, such diagnostic methods for community profiling were compared to reveal their reliability and their relation to classical clinical microbial diagnostics. Multiple factor analyses were performed to compare the different bacterial communities in a cohort of 56 adult CF patients and distinct polymicrobial consortia could be statistically identified and their persistence evaluated.

2.3 Material and Methods

2.3.1 Patient cohort and sample collection

Sputum samples were collected in sterile containers from 56 CF patients recruited in the CF outpatient clinic of the Hannover Medical School (MHH; Hannover, Germany), including a subgroup of 13 patients who provided sputum twice (n=10) or three times (n=3) within the sampling period of 2 years. In total, 72 separate sputum samples were collected after ethical approval for the current study was granted by the local health authority ethics committee. Sputum collection was done during the routine medical examination. The patients were between 18 - 51 years old and a median age of 31 was calculated for the cohort. Both genders were equally represented (48.2 % female and 51.8 % male). Signs of acute infections were observed for some patients, but most of them were in a stable phase of their disease and had no pulmonary exacerbation by the time of sputum collection. Sputum samples were collected in duplicates. Containers with sputum for DNA extraction were stored at -20°C. Sputum samples for clinical microbial diagnostics were immediately processed in the local microbiological laboratory of the MHH. Methods were applied according to the German

Quality Standards in Clinical Microbiology and Infectious Diseases (Hogardt 2006). Pulmonary functions were measured as forced expiratory volume in 1 second (FEV_1) with Ganshorn Body Scope version LF8.5E (Ganshorn Medizin Electronic; Niederlauer, Germany) and assessed according to the predicted normal values for adults defined by E.G.K.S. 1993 (The European Respiratory Society 1993). $FEV_1 \geq 90$ % is considered to be normal, however, some lung disease may be present. FEV_1 70 - 89 % shows mild lung malfunction, 40 - 69 % indicates moderate lung malfunction and less than 40 % is a sign of severe lung malfunction (Cystic Fibrosis Foundation Patient Registry 2012).

2.3.2 Sputum preparation and DNA extraction

An optimized protocol for the extraction of fungal and bacterial DNA was developed by modifying the manufacture's instructions of the kit and by adaptation of guidelines for the preparation of sputum samples recommended for detection of *Mycobacterium tuberculosis* in sputum (ifp Institut für Produktqualität 2008). Briefly, sputum samples were aliquoted and boiled for 15 minutes. To decrease the viscosity, a cysteine buffer (2 % NaOH, 1.45 % sodium citrate and 0.5 % N-acetylcysteine) was added in the same volume and the solution was mixed for 40 minutes. Milli-Q water was added to a final volume of 15 ml and centrifuged for 30 minutes at 4000 g. Supernatant was discarded and the pellet resuspended in 300 µl lysis buffer (20 mM Tris-Cl with pH 8.0, 2 mM sodium EDTA, 1.2 % Triton® X-100). 6 mg of Lysozyme (SERVA; Heidelberg, Germany) was added and incubated for 30 min at 37°C. To the total volume of the solution 0.5 % of β-mercaptoethanol was given together with 50 U of Lyticase (Sigma-Aldrich; Germany). The solution was incubated for another 45 min at 37°C before centrifugation at 12.000 rpm for 10 min. The pellet was resuspended in 300 µl Lyse T buffer from GeneMATRIX Tissue & Bacteria DNA purification kit (EurX Roboklon; Berlin, Germany) and 20 µl of Proteinase K (Qiagen; Hilden, Germany) was added. After incubation of 2 hours at 56°C, the DNA was extracted by following the instructions of the manufactures. DNA extracts were kept frozen at -20°C for further analysis.

2.3.3 Fingerprints by SSCP electrophoresis and sequencing of individual bands

PCR amplification of parts of the rRNA gene from bacteria was performed, using the following primers 27F (5´-AGAGTTTGATCMTGGCTCAG-3') and 521R (5'-ACCGTGGCTGCTGGCAC-3'). Amplicons of the 16S rRNA gene included the variable regions V1-V3 of the small subunit of the ribosome (SSU) with a amplicon size between 459 bp for *Nocardia sp.* and 505 bp for *Veillonella sp.* (Wos-Oxley et al. 2010). PCR was carried

out using 50 ng DNA of sputum extractions in a final volume of 50 µl, starting with an initial denaturation for 15 min at 95°C. A total of 30 cycles (1 min at 95°C, 40 sec at 56°C, and 1 min at 72°C) was followed by a final elongation for 10 min at 72°C. 1.5 U of HotStarTaq DNA polymerase was used for all amplifications (Qiagen). For single-strand DNA (ssDNA) preparation, reverse primer 521R was 5´-biotin labeled and magnetic streptavidin coated beads (Promega, Madison, Wis.) were applied to obtain ssDNA from the PCR amplicons according to Eichler et al. (Eichler et al. 2006). Dried pellets of ssDNA were resuspended in 7 µl of gel loading buffer (95 % formamide, 10 mM NaOH, 0.25 % bromphenol blue, 0.25 % xylene cyanol). After incubation for 10 min at 95°C, the ssDNA samples were shortly stored on ice and loaded onto a non-denaturing polyacrylamide-like gel (0.6x MDE gel solution; Cambrex BioScience, Rockland, Maine) for SSCP electrophoresis. Molecular fingerprints of bacterial DNA were obtained at 20°C and 400 Volt for 21 hours. More details on the SSCP fingerprints are given in Eichler et al. (Eichler et al. 2006). The gel was silver stained according to the method described by Bassam et al. (Bassam et al. 1991). SSCP gels were digitized using a HP Scanjet G4050 scanner. Bands with intensities of ≥5% of the total lane were considered for further statistical analysis. Intensities were determined by using the GelCompare II software (Applied Maths, Kortrijk, Belgium). Bands were excised from the SSCP gel, boiled in Tris buffer (10 mM Tris-HCl, 5 mM MgCl2, 5 mM KCl, 0.1 % Triton X-100, pH 9.0) and reamplified according to the conditions described above. Purified amplicons were sequenced by cycle sequencing (ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit; Applied Biosystems, Foster City, Calif.), products were purified again with the BigDye Terminator purification kit (Qiagen) and analyzed using capillary electrophoresis with fluorescence detection (ABI Prism 3100 Genetic Analyzer).

2.3.4 Preparation of Illumina high-throughput sequencing

PCR amplification of parts of the 16S rRNA gene, including the variable region V3, was performed by using the following primers 341F (5´-ATTACCGCGGCTGCTGG-3') and 518R (5'-CCTACGGGAGGCAGCAG-3') (Bartram et al. 2011). Amplicon library preparation was performed according to Camarinha-Silva et al. (Amélia Camarinha Silva 2011). Briefly, the 16S rRNA primers were designed with integrated complementary sequences to the Illumina specific adaptors to the 5'ends (**Table S1**). Further, a 6 nucleotide error correcting barcode (BC) was integrated in the forward primer, together with a 2 nucleotide CA linker to avoid amplification bias (Hamady et al. 2008) (Meyer & Kircher 2010). PCR was carried out using 50 ng DNA of sputum extracts in a final volume of 50 µl, starting with an initial denaturation for 15 min at 95°C. A total of 20 cycles (10 sec at 98°C,

10 sec at 53°C, and 45 sec at 72°C) was followed by a final elongation for 2 min at 72°C. Specific Illumina multiplexing sequencing primers as well as index primers were integrated with accordingly designed primers in a second subsequently performed PCR reaction. The second PCR was carried out using 1 µl of the reaction mixture as template. Additionally, 5µl of Q solution (Qiagen) were added to the final volume of 50 µl. Initial denaturation for 15 min at 95°C was followed by a total of 20 cycles (10 sec at 98°C, 10 sec at 62°C, and 45 sec at 72°C) and a final elongation for 2 min at 72°C. PCR products were separated on a 2% agarose gel and bands of correct size (around 330 bp) extracted and recovered using the QIAquick Gel Extraction kit (Qiagen). For each index, equimolar ratios of amplicons (50 ng) were pooled, all tagged with unique barcodes. Quality of these libraries as well as the DNA concentration was measured by an Agilent Bioanalyzer. Libraries were sent for paired-end multiplex sequencing on a GAIIx Genome Analyzer. Image analysis and base calling were done by using the Illumina Pipeline (version 1.7).

2.3.5 Data processing and phylogenetic analysis

Phylogenetic analysis of sequences was performed with the NCBI Tool BLAST/blastn and Ribosomal Data Base Project (RDP) Seqmatch Tool (Cole et al. 2009). Closest taxonomical groups were determined by sequence similarity and height of ssDNA bands in the molecular fingerprints. For sequences obtained from SSCP fingerptints, species identification was defined for sequence similarity of ≥ 98 %, ssDNA bands of same heights were considered to be same species. Further, phylogenetic distance analyses were performed to confirm the identification of the individual operational taxonomic unit (OTU) (data not shown). For sequences obtained by Illumina sequencing, only the forward end sequence reads were processed. A quality filter was applied and determination of representative reads were both done according to Camarinha-Silva et al. (Amélia Camarinha Silva 2011). Only 90 bp of the forward 5' end of all reads were used for further analysis, due to decreasing sequence quality. Representative reads were defined as i) present in at least one sample in a relative abundance > 1 % of the total sequences of that sample, ii) present in at least 2 % of samples at a relative abundance > 0.1 % or iii) present in at least 5 % of samples. Median number of total reads per sample was 10,933 sequences. For individual samples which were prepared and analyzed more than once, only the one with the most total sequence-reads was taken into account for the comparison of all 72 sputum samples. Species identification was defined for sequence similarity of ≥ 98%. However, sequences with ambiguous results were defined according to the closest shared taxonomical level. Sequences affiliated to human DNA were removed

before calculation of relative abundances of bacterial OTUs. Results from different Illumina runs were pooled and a cut-off level of 0.5 % relative abundance was chosen for each sample, in order to analyse only OTUs with a minimum relevance for the community. In total 65 OTUs from 72 sputum samples were further processed (**Table S2**).

2.3.6 Statistical analysis

Multivariant statistical analyses were performed using PRIMER 6 (Version.6.1.6, PRIMER-E, Plymouth Marine Laboratory, UK). Resemblance matrices were calculated using the Bray-Curtis similarity coefficient (Bray & Curtis 1957). According to the similarities the samples were ordinated in non-metric multidimensional scaling (MDS) plots (with 50 random restarts), distances between samples in these plots reflect similarities of associated OTU compositions. The reliability of MDS plots is indicated by the 2D stress value, which basically measures the differences between ideal distances in higher dimensional space and the actual distances in two dimensional space. Hierarchical agglomerative clustering was performed for the resemblance matrices using the group-average mode and SIMPROF (similarity profile permutation test), which tests for significant groups in an *a priori* unstructured data set (permutations for mean profile: 1000; simulation permutations: 999). To compare two resemblance matrices, the RELATE routine analysis was applied. This comparative (Mantel-type) test measures the similarity and calculates the Spearman's rank correlation coefficient Rho.

2.4 Results

2.4.1 Comparisons of methods for microbial composition profiling

The bacterial community composition of 72 sputum samples was analyzed using Illumina sequencing (NGS) and molecular fingerprinting by SSCP electrophoresis (fingerprinting). In order to validate the reproducibility and accuracy of the results, all samples were analyzed separately with both methods. Parts of the 16S rRNA gene were amplified, comprising the V1-V3 region for fingerprinting and only the V3 region for NGS. Accordingly, OTUs were defined and identified for both methods individually and their frequency of occurrence within the 72 sputum samples compared (**Figure 1**). For this direct comparison of fingerprinting and NGS, abundant OTUs were defined as ≥ 5 % relative abundance. Overall, a high degree of agreement for the frequency of occurrence of individual OTU for both methods was observed. High occurrence rates in both methods of at least 10 % positive samples were detected for the

dominant genera *Pseudomonas*, *Staphylococcus*, *Streptococcus*, *Rothia* and *Prevotella*. For both methods, a similar diversity of OTUs with lower occurrence rates was observed likewise. Discrepancies between NGS and fingerprinting were detected in the genus *Streptococcus*. Four additional OTUs within the genus *Streptococcus* were defined only by fingerprinting and therefore biased the comparability of detected occurrence rates for both methods. The higher taxonomic resolution in fingerprinting was achieved due to longer fragments of the 16S rRNA gene, comprising three variable regions. Nevertheless, most OTUs defined by NGS could unambiguously be affiliated to the respective OTUs defined by fingerprinting.

Besides the occurrences, also the relative abundances of all OTUs were analyzed and the community compositions of all 72 samples compared. The community compositions were assessed by the relative abundance of each individual OTU within one sample. The relative abundance of individual OTUs in NGS was defined equal to the relative frequency of the associated bacterial sequences. Results from fingerprinting and NGS correlated strongly, though partly different OTUs were defined due to different taxonomical resolutions (**Figure S1)**. Thus, comparison of both compositional data sets revealed a rank correlation value (Rho) of 0.816 (P-value= 0.01), which indicates highly similar community compositions in the sputum samples revealed by both methods (**Table 1**). A histogram of this similarity analysis is shown in **Figure S2a**. In **Figure 2**, the high degree of agreement is illustrated by comparisons of the relative abundance of the most dominant OTU in each individual sample, in disregard of the taxonomy.

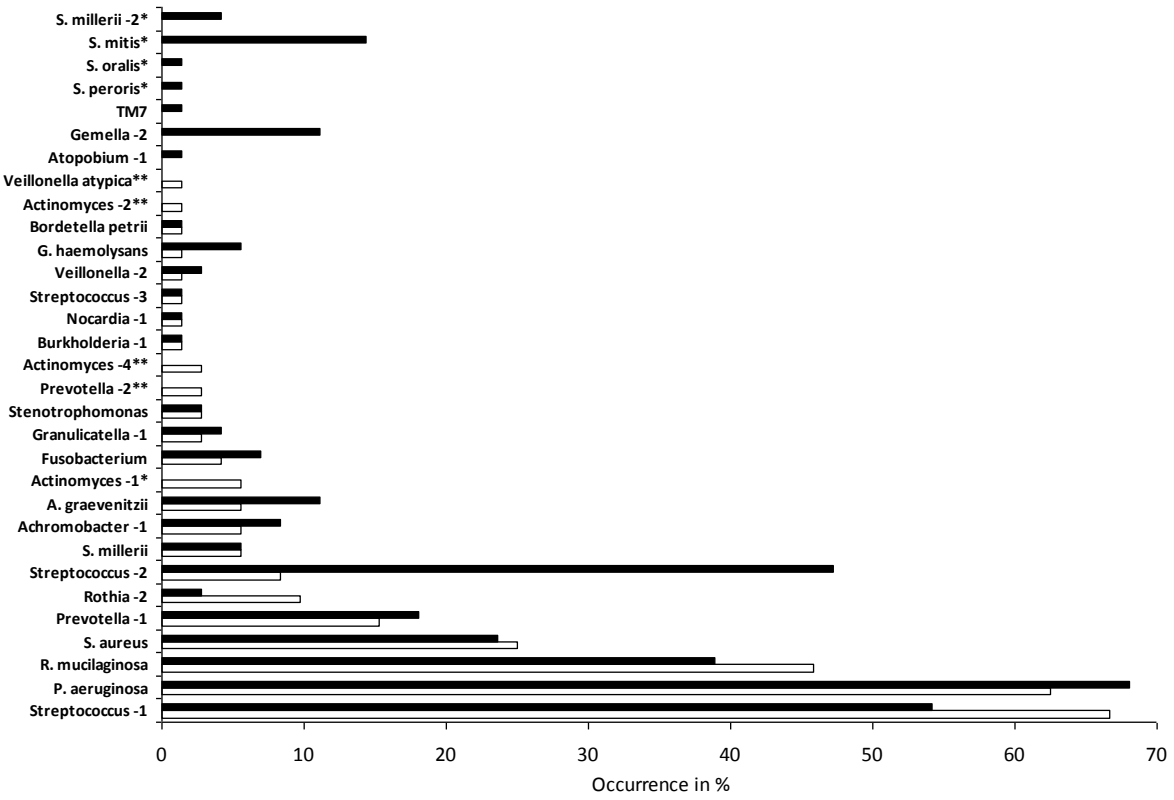
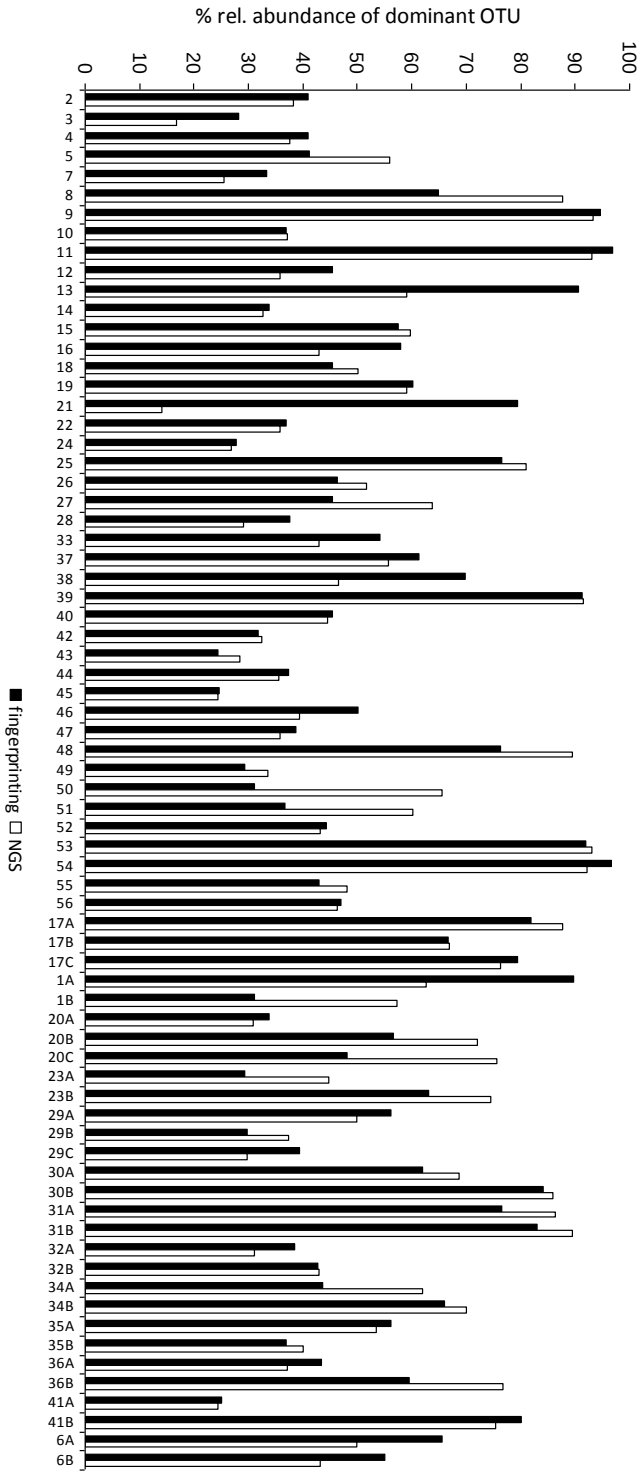


Figure 1: Occurrence of OTUs observed in 72 sputum samples with Illumina (white bars) and SSCP fingerprinting (black bars). OTUs are listed on the left according their occurrence observed with Illumina. OTUs defined only in SSCP fingerprinting are marked with * and those defined only in Illumina are marked with **. Only OTUs with ≥ 5 % relative abundance were taken into account.

Although a higher taxonomic resolution could be achieved by fingerprinting, NGS assured identification of OTUs at lower relative abundances. Whereas OTUs with ≥ 5 % relative abundance were considered for the comparisons of the different methods, microbial composition profiling based on NGS took OTUs into account with a relative frequency of ≥ 0.5 % if a minimum of 2,000 sequence-reads per sample was achieved. An overview of all OTUs observed by NGS with relative abundances of ≥ 0.5 % as well as their minimum and maximum relative abundance in individual samples is given in **Table S2.**

Figure 2 (below): Comparisons of the relative abundance of the most dominant OTU in each individual sample assessed with Illumina (white bars) and SSCP fingerprinting (black bars). Numbers indicate patients and letters are according order of sputum collection. Samples are listed according their order in the more detailed compositional comparisons shown in Figure S1.



Subsets of sputum samples were chosen to be analyzed repeatedly to estimate the extraction and amplification bias. From a subset of eight sputum samples, DNA extractions were performed twice from different aliquots. Sputum samples from this subset exhibited different viscosity and visual appearance as well as considerable differences to each other concerning the bacterial community composition. Comparison of the compositional data assessed by NGS from the first and second DNA extraction of each sample revealed a high degree of agreement (**Figure 3**). The same OTUs were detected in both extractions. Discrepancies were

observed in relative abundance of sequences but mostly moderate. The highest discrepancy was observed for the abundances of *Pseudomonas* -1 in the sample e/e*, followed by *Pseudomonas* -1 of the sample b/b*. Comparing all OTUs of corresponding DNA extractions in the subset, a mean discrepancy of 2.9 % relative abundance was calculated.

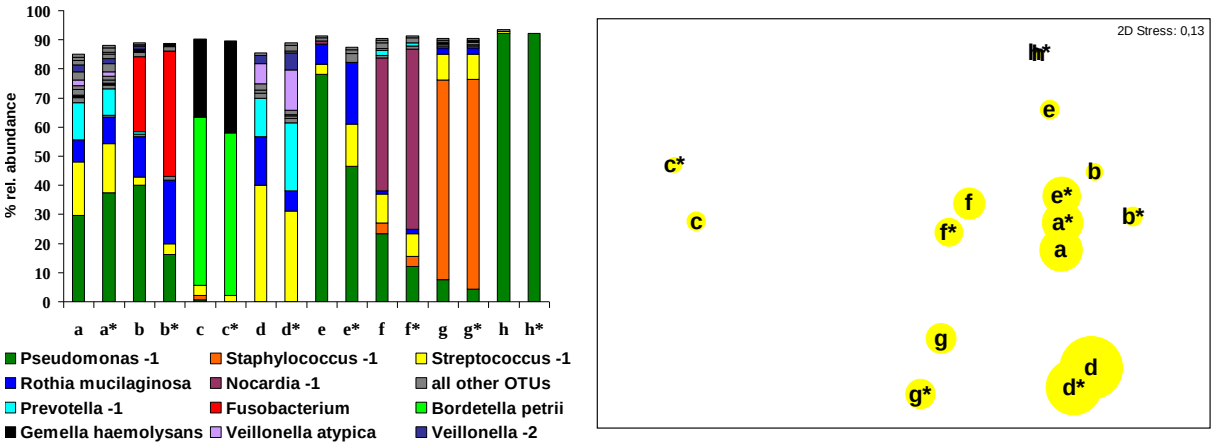


Figure 3: Comparisons of community composition of sputum samples (a – h). Two different DNA extractions per sample were analyzed. **Left:** Comparisons of relative abundances of bacteria. Colour code is given beneath the diagram. **Right:** Corresponding MDS plot to the diagram. Dissimilarity matrix was calculated with Bray-Curtis algorithm taken all bacterial abundances into account and displayed in the MDS plot. Exemplarily, relative abundances of *Streptococcus* -1 were indicated by bubbles, sizes are according the abundances (max. 40%). .2D Stress value of 0.13 indicated moderate stress on the plot.

A resemblance matrix was calculated including the first and second DNA extraction and samples were accordingly ordinated in a MDS plot (**Figure 3**). To further assess the extraction bias for the community composition of each sample. The more similar the community compositions of each sample, the more close they ordinated together in the MDS plot. To illustrate the coherence of the individual relative abundance of each OTU and the ordination of the samples in the MDS plot, the abundances of *Streptococcus* -1 were exemplarily highlighted for each sample in the figure. It is important to note here, that for the distances between samples the abundances of all bacteria were taken into account. The MDS plot revealed high similarity of community compositions in each corresponding DNA extraction. Again, the highest discrepancies were observed for sample e/e* and b/b*. Individually calculated resemblance matrices for the first and second DNA extraction were compared and revealed a rank correlation value (Rho) of 0.911 (P-value= 0.01), indicating strong agreement of the results (**Table 1**). The histogram of this similarity analysis is shown in **Figure S2b**.

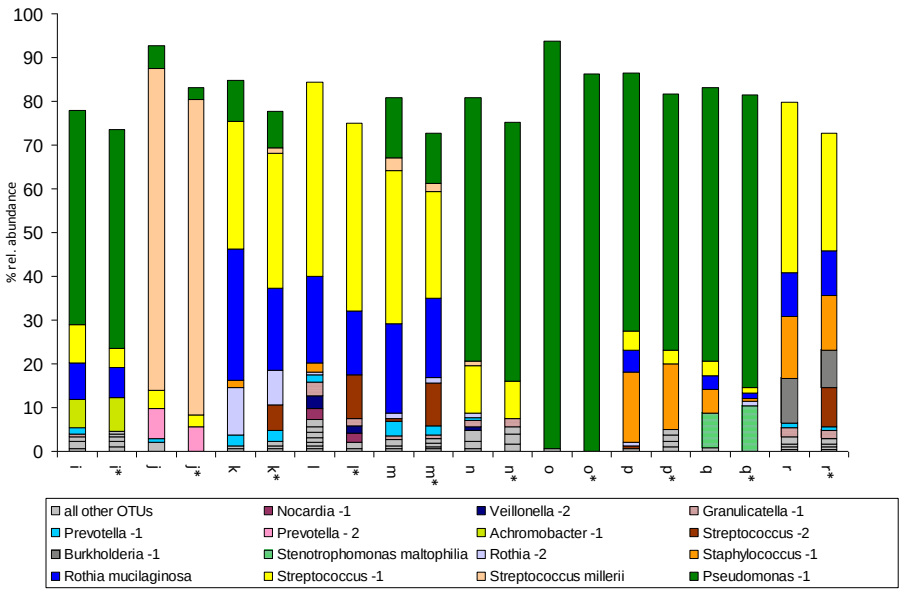


Figure 4: Comparisons of the community composition of sputum samples (i – r). Two different preparations of each sample were performed and separately analyzed with NGS. Samples are individually labelled, those with * were analyzed in the second run.

The amplification bias for NGS-based assessment of the community composition was estimated likewise. A subset of ten samples was analyzed twice. PCR amplicons and libraries were prepared separately and NGS performed in different runs. Compositional data are shown in **Figure 4**. Comparisons revealed a high degree of agreement, indicated by a rank correlation value (Rho) of 0.976 (P-value= 0.01). Comparing all OTUs of corresponding amplifications in the subset, a mean discrepancy of 3.5 % relative abundance was calculated (**Table 1**). The histogram of this similarity analysis is shown in **Figure S2c**.

Data sets compared	Spearman rho	Major discrepancies
Illumina; different extractions	0.911	mean of 2.9 % differences in OTU abundances
Illumina; different sequencing reactions	0.976	mean of 3.5 % differences in OTU abundances
Illumina and SSCP fingerprinting; all samples (≥ 5%)	0.816	different taxonomic resolution for the genus *Streptococcus*

Table 1: Comparison of data sets from bacterial community elucidation. Resemblance matrices were calculated for each dataset and compared using spearman rank correlation (rho). Major discrepancies observed for each pair of data sets are mentioned. For the comparison of SSCP fingerprinting and Illumina, only OTUs with a relative abundance of ≥ 5 % were considered.

The occurrence rates of these OTUs were additionally compared with the results of the clinical diagnostics: Simultaneously collected sputum samples from each patient were analyzed with conventional culturing methods in the microbiology department of the clinic. Results were reported and compared with the results of the NGS analyses (**Figure 5a**). High degree of agreement was achieved for the known pathogens, like *P. aeruginosa*, *S. aureus*,

B. cepacia, *A. xylosoxidans*, *S. maltophilia*. Most other species detected by NGS were not observed with culturing, including all strict anaerobic bacteria, like *Streptococcus millerii* or *Fusobacterium nucleatum*. Furthermore NGS revealed a considerably higher detection frequency of alpha-hemolytic *Streptococci* than culture-based diagnostics. Very few species were detected only once or twice by cultivation and not observed in NGS.

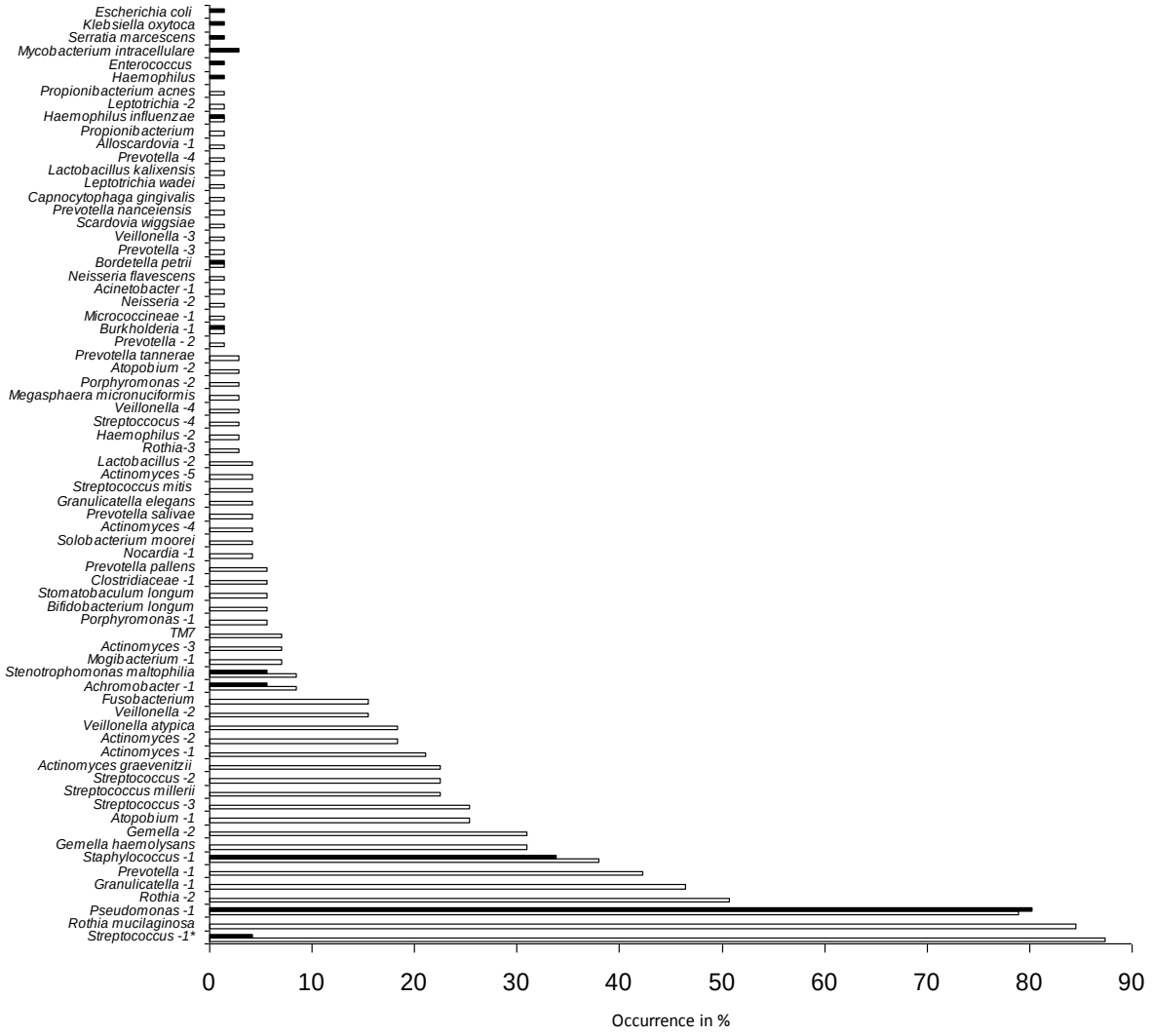
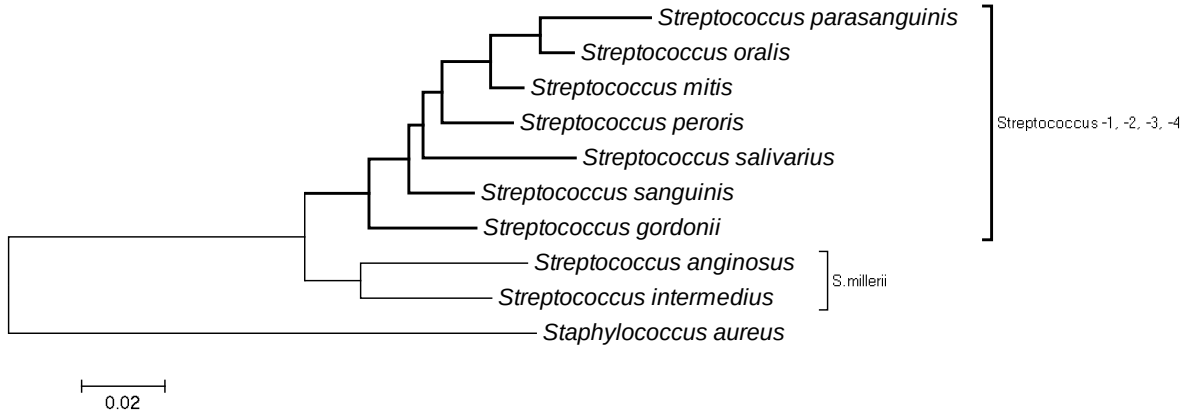


Figure 5a (above): Occurrence of OTUs observed in sputum samples analyzed with Illumina (white bars) and culturing for diagnostics (black bars). OTUs are listed on the left according their occurrence observed with Illumina. Only 71 sputum samples were taken into account (no reported data from culturing for sample #20B). * Streptococcus-1 was compared with alpha-hemolytic Streptococci.

Figure 5b (below): Phylogeny of detected *Streptococcus* species in sputum samples. High sequence variation was observed with SSCP fingerprinting. Closest described representatives from databases are given in the taxonomic tree. Associated OTUs defined in NGS are indicated with brackets. Aerobic alpha-hemolytic Streptococci are further marked with bold branch lines. *Staphylococcus aureus* was included for rooting of the tree. Scale bar represents base substitution per site.



Since the greatest differences between NGS and clinical diagnostics were observed for the genus *Streptoccus*, phylogeny of respective species in sputum samples was further elucidated **(Figure 5b)**. Sequences from SSCP fingerprinting revealed several aerobic alpha-hemolytic *Streptococci*. The individual relative abundances of these species in sputum samples varied strongly, but *S. parasanguinis* and *S. salivarius* were the dominant bacteria in several samples (**Figure S1**). Additionally two distinct species of the strict anaerobic *S. millerii* group were identified.

2.4.2 Community composition and dominant bacterial species

Although some differences between compositional analyses with SSCP fingerprinting and NGS were observed, the overall similarity was shown to be high and with a good resolution of the community composition by NGS. Therefore, results from NGS were taken for further microbial community profiling. Higher taxonomic resolution for individual OTUs in NGS could be achieved by comparisons with the strongly correlating results obtained by fingerprinting. Although sequence similarity of NGS alone would not assure the species identification, if OTUs defined in NGS were unambiguously and equally assigned with both methods they could be affiliated to the species level according to the fingerprinting results (**Table S2**). Accordingly, all 72 sputum samples were analyzed, the compositional similarities determined and samples ordinated in one MDS plot (**Figure 6a**). Although individual bacterial community compositions of each sample were partially highly divers, all of them exhibited at least one of the previously identified dominant OTUs *Pseudomonas* -1, *Staphylococcus* -1, *R. mucilaginosa* and *Streptococcus* -1 (**Figure 6b**). A distinct distribution of sputum samples according the abundances of these OTUs was shown and four categories were observed: The major CF pathogens *P. aeruginosa* and *S. aureus*, represented by the accordingly affiliated OTUs *Pseudomonas* -1 and *Staphylococcus* -1, characterized the first two categories, whereas category three and four was characterized by the abundances of *R.*

mucilaginosa and *Streptococcus* -1 respectively. The OTU *Streptococcus* -1 could not be affiliated to one distinct species but exhibited high sequence similarity to *S. salivarius* and *S. parasangunis*. Together with *R. mucilaginosa*, both are not considered to be classical pulmonary pathogens in CF. However, no explicit borders between those categories could be defined. Samples with high relative abundances of these dominant OTUs were ordinated in the individual periphery of each category. For *P. aeruginosa* and *S. aureus* communities with respective individual relative abundances of ≥ 80% were observed, whereas the maximum relative abundance of *R. mucilaginosa* and *Streptococcus* -1 was 52% and 50% respectively. Samples with moderate abundances of the individual dominant OTU were more likely ordinated in the centre of the plot. Overlaps were observed for all four categories. Furthermore, the plot revealed some outlier. Especially samples #5, #20B, #20C, #34A, #35A, #2 and #16 were ordinated differently in comparison to the other samples, demonstrating rather different community compositions observed in these samples.

Fig. 6a

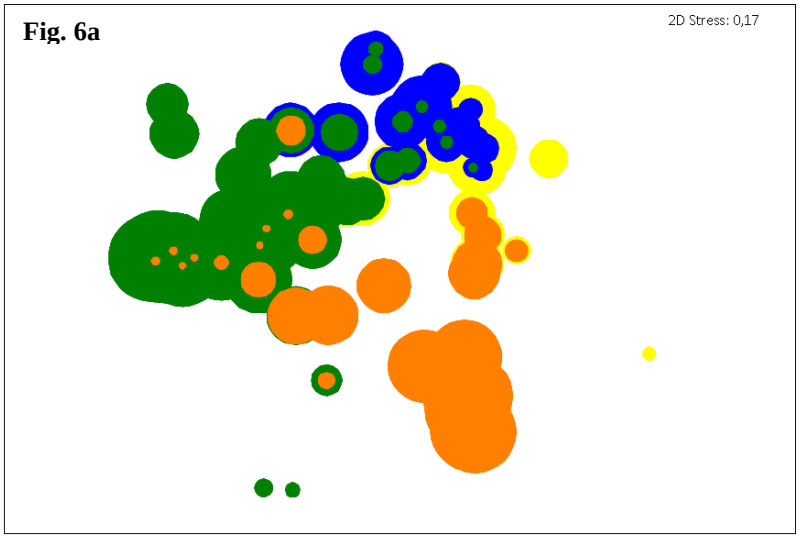


Fig. 6b

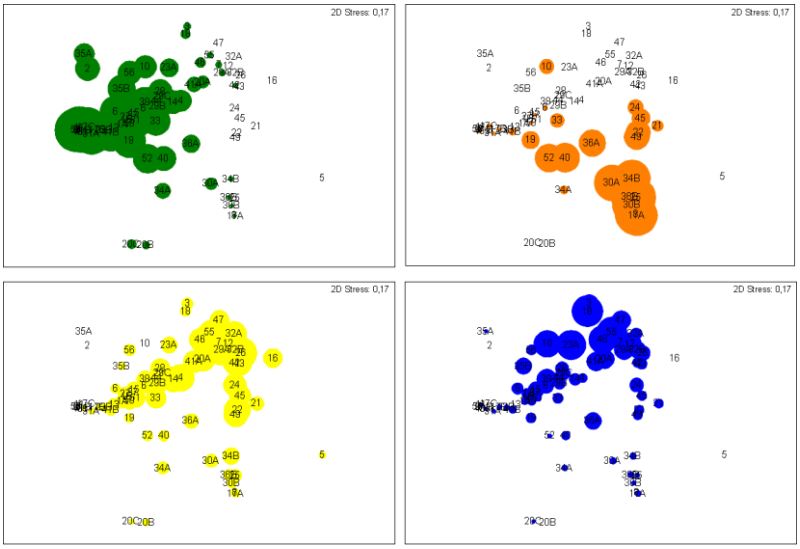


Figure 6a, b: MDS plots from bacterial community composition observed in 72 sputum samples. Bubbles indicate relative abundances of dominant OTUs in the cohort: *P. aeruginosa* in green, *S. aureus* in orange, *Streptococcus* -1 in yellow, *R. mucilaginosa* in blue. 2D Stress values are given in each plot and reveal moderate stress. **6a:** Merged bubble plots from all four dominat OTUs. Coloured bubbles indicating the individual abundances. *P. aeruginosa* as well as *S. aureus* have priority. Abundances for *R. mucilaginosa* and *Streptococcus* -1 are in background. **6b:** Individual MDS plots for each dominant OTU. Numbers indicate individual patients. Letters are according order of sputum collection.

2.4.3 Comparison of host factors and the bacterial community composition

The relationships between bacterial communities, host factors and clinical aspects were analyzed. Comparisons of factors and aspects between samples from all four categories were of particular interest. In order to examine any correlation between the overall community composition and the patient's age, the cohort was divided into four age classes according to the corresponding statistical quartils. Symbols for each age class were superimposed in the previous MDS plot calculated for the community composition of all 72 sputum samples (**Figure 7a**). No correlation was observed between age of the patient and the four previously defined categories of bacterial communities. Patients with a community composition dominated by the major CF pathogens *P. aeruginosa* and *S. aureus* were distributed equally in all age classes as well as patients with high relative abundances of *R. mucilaginosa* and *Streptococcus* -1. In more detail, the relationship of *P. aeruginosa* and age is demonstrated (**Figure S4**). Samples were ordered by the age of patients at the time point of sputum collection and compared with the individual relative abundances of *Pseudomonas* -1. Again, no correlation was observed. Although most samples without presence or with low relative abundances (≤ 5 %) of *P. aeruginosa* were observed for patients with ≤ 25 years, similarly low abundances were also detected in several samples of other age classes. Likewise samples from patients of all ages exhibit communities dominated by *P. aeruginosa*. In summary, a tendency for young adult CF patients to exhibit less frequently communities with highly abundant *P. aeruginosa* could be shown but exceptions in individual cases were observed.

In order to examine any correlation between the overall community composition and the lung function of the patients, the cohort was divided into four classes according to the statistical quartils of the FEV_1 values of all patients. Symbols for each class were superimposed in the previous MDS plot calculated for the community composition of all 72 sputum samples (**Figure 7b**). A median predicted FEV_1 of 35 % was calculated for the cohort and patients with strongly reduced lung functions (≤ 25 %) were observed in all four previously defined categories of bacterial communities. Likewise, relatively high FEV_1 values of ≥ 51 % were observed for patients with communities dominated by the major CF pathogens *P. aeruginosa* and *S. aureus*. Further demonstrated were the relationship between *P. aeruginosa* and the lung function: Samples were ordered by relative abundances of *Pseudomonas* -1 and compared with the predicted FEV_1 values (**Figure 8**). No clear correlation was observed between the lung function and the relative abundance of *P. aeruginosa*. Like described above, although all patients with FEV_1 values of ≤ 60 % exhibited none or low relative abundance of *P.*

aeruginosa in sputum samples, relatively good lung functions (≤ 51 %) were also measured for some patients with *P. aeruginosa*-dominated bacterial communities. Likewise, strongly reduced lung functions of ≤ 25 % and no or low relative abundance of *P. aeruginosa* was observed for the same patients.

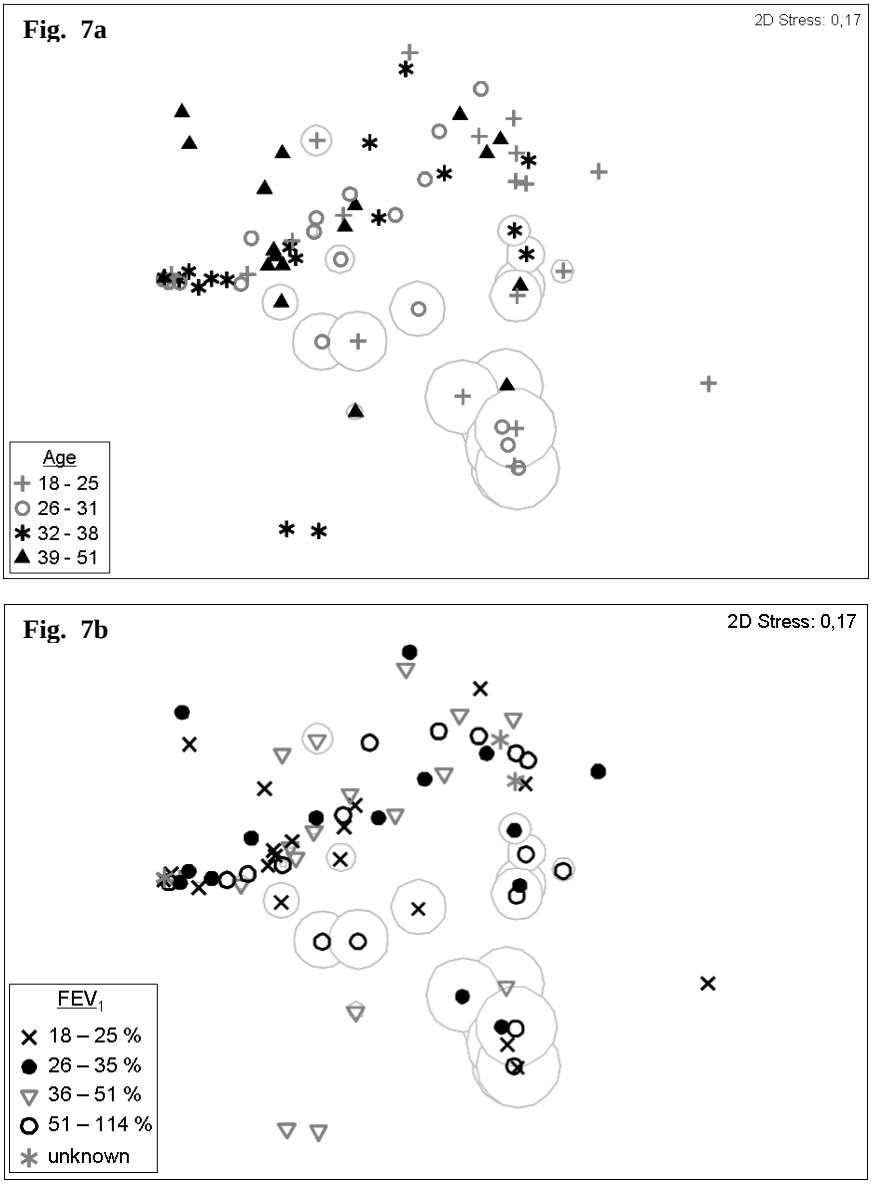


Figure 7a, b: MDS plots from bacterial community composition observed in 72 sputum samples with superimposed host and clinical factors. For a better orientation in the plot, relative abundances of *S. aureus* are indicated like previously shown. **7a:** Age of patients superimposed on the previous MDS plot calculated for the community compositions of 72 sputum samples. Four age classes were determined for the cohort according to the statistical quartils. Symbols indicate the age of the patient at the time point of sputum collection. **7b**: Lung function of patients superimposed on the previous MDS plot calculated for the community compositions of 72 sputum samples. Four classes of lung functions were calculated according to the statistical quartils of FEV_1 values measured for the patients on the day of sputum collection. Symbols indicate the four classes and samples without associated FEV_1 values.

Correlation analyses were further performed between the genotype of the CFTR gene of each patient and the corresponding community composition in the sputum samples (**Figure S5**). Patients were classified according the genotypes: 41 % were homozygous-ΔF508, 16 % heterozygous-ΔF508, 16 % had a non-ΔF508 mutation and for 27 % of the patients the precise genotype was not or only partly determined. Neither of the genotypes correlated

significantly with communities dominated by *P. aeruginosa*, *S. aureus*, *R. mucilaginosa or Streptoccocus* -1. However, the class of patients with non-ΔF508 mutations was composed of diverse genotypes. In only three samples from patients out of this class, the community was dominated by *P. aerugionsa* with relative abundances ≥ 50 % and in seven samples the communities exhibited only low abundance of *P. aeruginosa* (≤ 5 %).

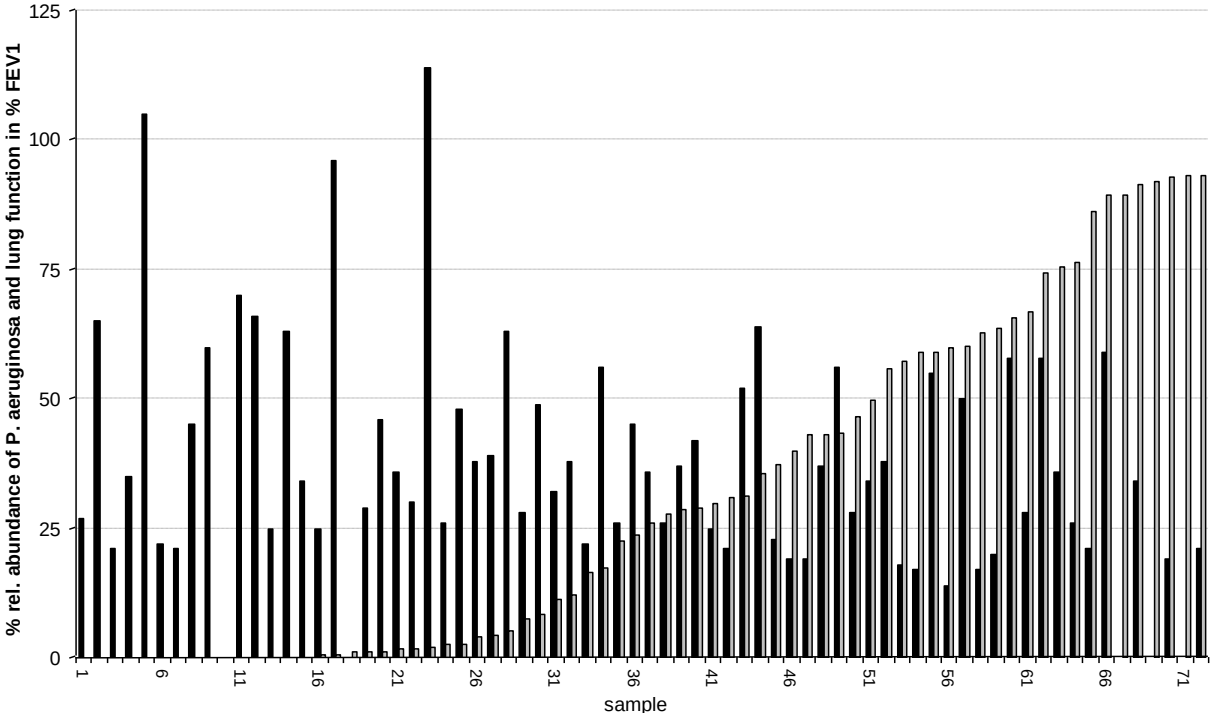


Figure 8: Comparison between relative abundance of *P. aeruginosa* (grey bars) and lung function of the patient (black bars). Samples are ordered by relative abundance of *P. aeruginosa*. Lung functions of the patients at the individual time point of sputum collection is measured by the predicted FEV_1 value. Both, relative abundance of *P. aeruginosa* and lung function are given in % on y-axis. For each patient, both parameters are indicated. Missing bars for *P. aeruginosa* indicate its absence in the sample and missing bars for FEV_1 indicates no measurement for the patient at time point of sputum collection. A median FEV_1 of 35 % was calculated for the cohort.

Overall, the high relative abundances of *P. aeruginosa*, *S. aureus, R. mucilaginosa* and *Streptoccocus* -1 in sputum samples could not be clearly correlated with any clinical parameter or host factor of the CF patients.

2.4.4 Community composition and polymicrobial consortia

To determine microbial subgroups within the four previously defined categories of bacterial communities, cluster analyses were performed with the community compositions from all 72 sputum samples. Further statistical determination of discrete clusters was performed by

SIMPROF. A low significance level of 1 % was applied and revealed 14 subgroups of sputum samples with distinct bacterial community compositions (**Figure 9**). These SIMPROF groups (A-N) were superimposed on the MDS plot previously calculated for all 72 sputum samples (**Figure 10**).

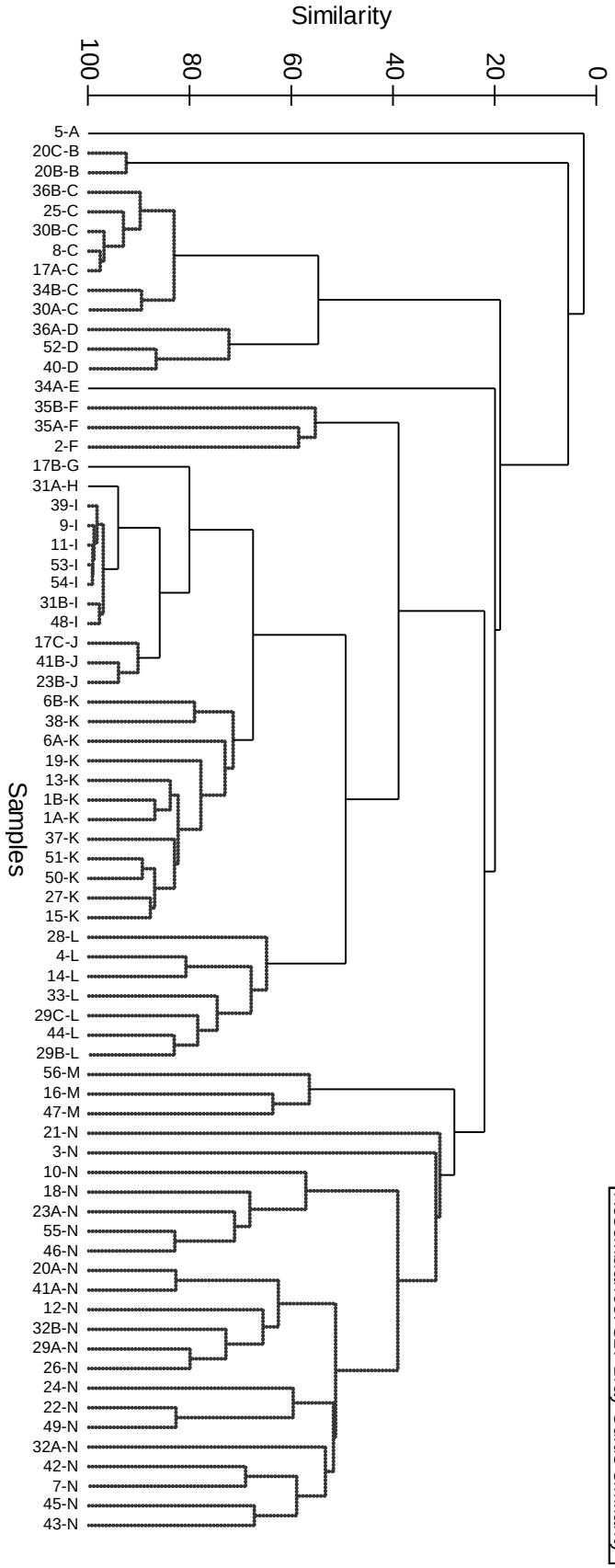


Figure 9 (left): Cluster analysis with SIMPROF test on bacterial communities observed in 72 sputum samples. Clustering was performed on the resemblance matrix based on Bray-Curtis similarities and in group average mode. A significance level of 1 % was applied for statistical determination of discrete clusters by SIMPROF. Numbers represent patients and letters represent the order of sputum collection (e.g. 20A, 20B, 20C). Hyphenated letters indicate SIMPROF groups (A-N). Discrete clusters in the dendogram are additionally highlighted with bold dashed lines.

.

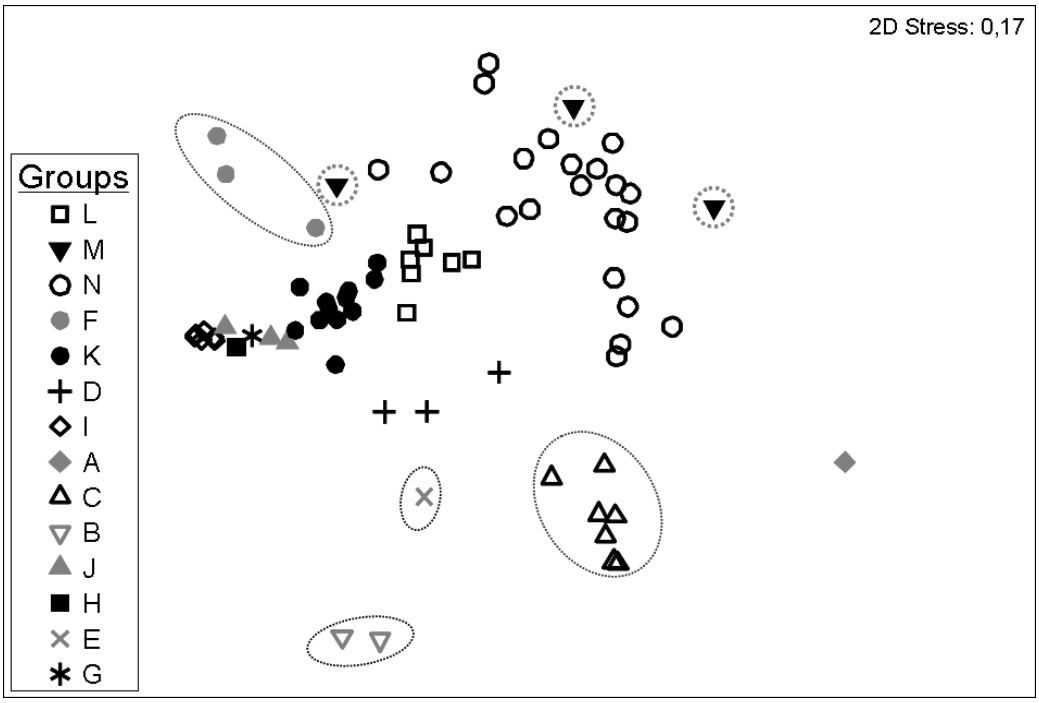


Figure 10: SIMPROF groups superimposed on the previous MDS plot calculated for the community compositions of 72 sputum samples. Distinct microbial communities determined by cluster analysis and SIMPROF test, are indicated by the individual groups A-N with associated symbols. All three samples associated with group M are marked with bold dashed circles and also samples of groups B, C, E and F are circled additionally for orientation

The plot demonstrated several distinct microbial communities within samples dominated by the major CF pathogens *P. aeruginosa* and *S. aureus*. Furthermore, previously observed outliers with distinct bacterial communities were accordingly determined as distinct groups: #5 in -A, #20B and #20C in -B, #34A in E, #35A and #2 were grouped together with #35B in -F, #16 was grouped with #47 and #56 in group -M. Whereas samples of most SIMPROF groups are ordinated closely together in the MDS plot, group M was more unequally distributed though all three samples of this group could still be considered within the previously defined categories of *R. mucilaginosa* and *Streptococcus* -1.

For each SIMPROF group, the characteristic abundances of bacterial species were observed and number of positive samples and patients were mentioned (**Table 2**). Characteristic bacteria were defined as those present in all associated samples. One exception was group N which consisted of more diverse samples in terms of microbial community composition. In group N, *R. mucilaginosa* was abundant in all samples and detected together with several anaerobic bacteria at high relative abundances in individual samples. Numbers of positive samples are mentioned for each species in table 2. Notably, SIMPROF group N was almost completely represented by samples dominated by *R. mucilaginosa* and *Streptoccocus* -1. For

all SIMPROF groups, ranges of relative abundances of each OTU were given. *P. aeruginosa* was defined characteristic for most of the groups (particularly in D-L). Groups I and J were differentiated only by different relative abundances, though both communities were strongly dominated by *P. aeruginosa*. All other SIMPROF groups were characterized by the consistent abundance of more than one bacterial species. Relative abundances of more than 50 % were only observed for the major CF pathogens *P. aeruginosa* and *S. aureus* as well as for *S. millerii*, *Fusobacterium nucleatum, Bordetella petrii* and *Nocardia sp.*.

Table 2	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**
P. aeruginosa		**3-4**		**17-43**	**12**	**17-43**	**67**	**86**	**89-93**	**74-76**	**43-66**	**24-36**		
Streptococcus -1			**2-14**								**4-19**	**18-38**	**19-20**	
R. mucilaginosa												**4-9**		**4-48** (n=21)
S. millerii		**72-76**												
S. aureus			**69-88**	**37-44**										
A. xylosoxidans													**36-48**	
Granulicatella -1												**1-3**		**0-6** (n=15)
S. maltophilia							**10**	**3**						
Prevotella - 2		**6-12**												
Nocardia -1					**62**									
Prevotella -1						**1-38**								**0-23** (n=15)
A. graevenitzii														**0-15** (n=10)
Veillonella atypica														**0-14** (n=9)
G. haemolysans	**31**													**0-2** (n=7)
Gemella -2												**1-4**		**0-2** (n=9)
F. nucleatum						**16-54**								
Bordetella petrii	**56**													
Actinomyces -1														**0-14** (n=9)
pos. samples	1	2	7	3	1	3	1	1	7	3	12	7	3	21
pos. patients	1	1	6	3	1	2	1	1	7	3	10	6	3	20
range % FEV_1	22	38 - 48	25 - 96	22 - 56	38	19 - 26	28	21	19 - 59	26 - 58	14 - 58	19 - 64	21 - 37	21 - 114

Table 2: Groups of bacterial communities according cluster analysis and SIMPROF test. Fourteen SIMPROF groups A-N are determined, numbers of positive samples and patients are shown. Groups are characterized by OTUs with consistent presence. Ranges of individual abundances of these characteristic OTUs are given in % for each associated group. Group N contains more diverse samples and characteristic bacteria are not consistently present in all samples, therefore the number of positive samples for each identified characteristic OTU in this group are given in brackets

Comparison between SIMPROF groups and lung function values, revealed consistently strongly reduced values for patients of group F which were characterised by highly abundant

Fusobacteria. Other groups comprised of the previously observed outlier in the MDS plot were also associated with strongly reduced lung functions, like group A characterized by co-occurring high abundant *G. haemolysans* and *B. petrii* and group M characterized by high abundant *Achromobacter xylosoxidans*. Further, associated with low lung functions were groups H-J with communities strongly dominated by *P. aeruginosa*, though exceptions of patients with moderate lung function were also observed ($FEV_1 \geq 51$ %). For the other groups, including group C and D both characterized by the highly abundant pathogen *S. aureus*, no such association was observed.

2.4.5 Community dynamics

Cluster analysis and SIMPROF revealed rather distinct bacterial communities partly characterized by infrequent bacterial species within the 72 analyzed sputum samples. To further examine the clinical relevance of the communities and of associated infrequent bacterial species, dynamics of the community composition were observed in a subcohort of 13 CF patients (**Table 3**). Sputum samples were collected within 4 month, 8 month or with more than 12 month to the previous sample from the same patient. Dynamics of bacterial communities were indicated by the SIMPROF group mentioned for each sample. Constant and changing community compositions were observed for sputum samples taken in all three time ranges. Strong changes within a short time were detected for patient #41, whereas constant communities were observed for patients #6 and #1 in a long time range of more than 12 month, both were constantly clustered into group K.

To examine the persistence of infrequent bacteria, the dynamics of communities characterized by these species were of particular interest. The communities over time were demonstrated for the patients #20, #34 and #35 (**Figure 11**). For patient #20, after a complete change of the community composition within the initial 12 month the community stayed constant with strong dominance by *S. millerii* in both samples (group B). Group F was associated with strongly reduced lung functions and characterized by highly abundant *Fusobacterium nucleatum*. In two associated samples of patient #35, communities stayed relatively constant over a period of 8 month. *Nocardia* -1 was detected with high relative abundance in patient #34. Though *Nocardia* -1 was still detectable in low relative abundance, a change of the community composition was observed within 8 month from SIMPROF group E to the *S. aureus*-dominated group C.

time	patient	group
	23A	N
> 12 month	23B	J
	31A	H
< 4 month	31B	I
	34A	E
< 8 month	34B	C
	36A	D
< 8 month	36B	C
	41A	N
< 4 month	41B	J
	17A	C
> 12 month	17B	G
< 8 month	17C	J
	20A	N
> 12 month	20B	**B**
< 4 month	20C	**B**
	29A	N
> 12 month	29B	**L**
< 8 month	29C	**L**
	6A	**K**
> 12 month	6B	**K**
	1A	**K**
> 12 month	1B	**K**
	32A	**N**
< 4 month	32B	**N**
	30A	**C**
< 4 month	30B	**C**
	35A	**F**
< 8 month	35B	**F**

Table 3 (left): Table shows subgroup of 13 patients giving sputum samples twice or three times. Time indicates the temporal distance to the previous sample. SIMPROF groups from cluster analysis of community compositions are mentioned. Constant communities in sequent samples of one patient indicated by constant SIMPROF groups are bold. Patients with constant microbial communities are in the lower part of the table. Dynamics of community composition. CF patients are given by numbers and sputum sample in chronological order by capital letters (A-C). Labels are according to the previous MDS plot with all 72 samples.

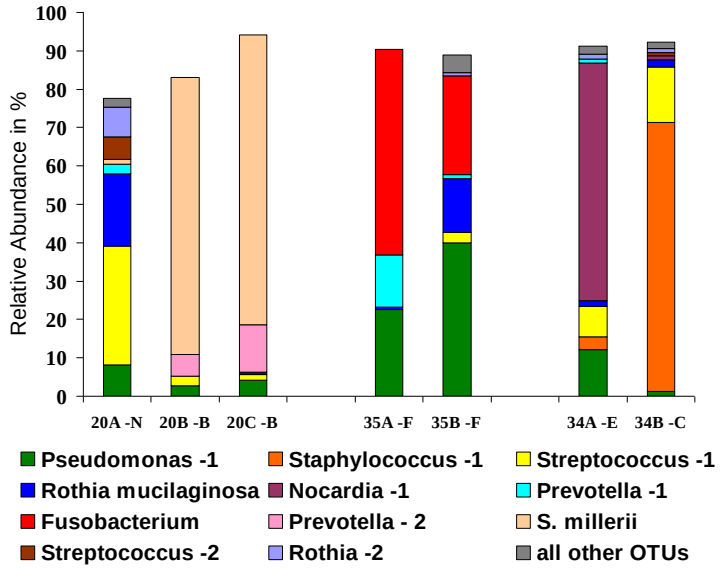


Figure 11 : Column plot shows dynamics of community composition in 3 individuals from subgroup of patients that gave samples twice or three times. Patients were further characterized by infrequent sputum communities revealed by cluster and MDS analysis. SIMPROF groups are mentioned with the according hyphenated letters. Each OTU is indicated by a specific colour and further defined in the legend. Dynamics of community composition. CF patients are given by numbers and sputum sample in chronological order by capital letters (A-C). Labels are according to the previous MDS plot with all 72 samples.

2.5 Discussion

2.5.1 Validation of methods for microbial community profiling

Elucidation of microbial communities with culture-independent methods has become a widely used approach in the past years and different methods were compared in order to validate their accuracy (Hori et al. 2006) (Caporaso et al. 2012) (Cleary et al. 2012). In the current study, conventional SSCP fingerprinting was compared with sequencing of barcoded amplicons using an Illumina short read genome analyzer. Good reproducibility of the results was observed: Community structures and compositions revealed by both methods are highly similar. However, some differences were detected but mainly due to different taxonomic resolutions. Using three variable regions in SSCP fingerprinting for affiliation of the different OTUs provided confident identification of the species, whereas the short reads used for Illumina provide reliable classification to the genus level in most cases. However, even short reads of the 16S rRNA gene are known to have good taxonomic information and enable microbial community profiling if primers are thoroughly chosen (Soergel et al. 2012). Likewise, a minimum read count of 2,000 sequences per sample chosen in the current study was also shown to be sufficient to recapture the bacterial community (Caporaso et al. 2011). Considering the low costs and rapid availability of results, SSCP fingerprinting can be used as an alternative to examine the compositional variation of microbial communities.

DNA extraction and amplification are crucial steps in sample preparation for culture-independent methods. Uneven spatial distribution for certain bacteria in biofilms are known and may also be found in sputum (Stewart et al. 1997) (Goddard et al. 2012). Therefore, especially DNA extraction from different aliquots of the sputum sample may bias the community profile. Respective comparisons of subsets with separately prepared samples revealed high conformity of the community compositions for both crucial steps. These comparisons suggest that bias due to DNA extraction and amplification only had minor influence and did not significantly alter the profiling of the bacterial communities in the CF cohort. Furthermore, these comparisons revealed that MDS plots represent a robust statistical analysis to display the similarity between samples that reflect minor changes in the individual compositions without overly affecting the total display of all the samples.

Clinical diagnostic of respiratory infections in adult CF patients is based on the detection of single pathogenic species by conventional culturing. Respective comparisons with NGS revealed high degree of agreement for well known CF pathogens like *S. aureus*, *P. aeruginosa*

and other *Proteobacteria*. However, great discrepancies were observed for the detection of species from the genus *Streptococcus*. Clinically relevant alpha-hemolytic *Streptocci* were under-diagnosed in comparisons to culture-independent methods. Correct identification of these bacteria by phenotypic methods, in particular the *S. mitis* and *S. sanguinis* group, is known to be difficult whereas sequencing methods allow at least identification of the correct streptococcal group (Haanperä et al. 2007). Anaerobic species, including the *S. millerii* group or *Fusobacteria*, were not detected at all by culturing but occurrence frequencies over 15 % in this cohort indicate a significant prevalence of these bacteria in CF patients. There clinical relevance is discussed below. Overall, culturing did not reflect the complex bacterial communities revealed by NGS.

Considering the growing interest in community analysis for CF, culture-independent methods are becoming more important (Rogers et al. 2009) (VanDevanter & LiPuma 2012). Fingerprinting methods allow confident elucidation of abundant bacteria, whereas the detection limit of NGS includes also rare species but might have less taxonomic resolution depending on the technique (Pedrós-Alió 2012). Whereas, occurrence frequencies of known CF pathogens, including *S. maltophilia* or *A. xylosoxidans*, were similar to the reported numbers for adult CF patients, in particular the detection of emerging pathogens will undoubtedly benefit from these methods (Bittar et al. 2008) (Hauser et al. 2011). Intuitively accessible statistical methods, like MDS plots, are essential to display compositional datasets and will further increase the knowledge about respiratory bacteria as integrated communities.

2.5.2 Single bacterial species versus polymicrobial consortia

Understanding the microbial flora is of considerable importance for treatment strategies of CF patients and inter-species interactions have to be considered to elucidate the precise contribution of different bacteria to patient morbidity (Harrison 2007). In the current study, a direct comparison of the microbial communities observed throughout the cohort was approached to understand their individual impact on health of the patients. Therefore, two strategies were tested and correlation with clinical parameter drawn, i) categorization of communities by single pathogens and ii) elucidation of communities by polymicrobial consortia.

In regard of the routine clinical diagnostics, the categorization of communities by single pathogens can be considered to be more conventional. Noticeable, the four most frequently

found OTUs also exhibited high relative abundances within individual samples and therefore defined the four categories of microbial communities. Such a correlation between frequency and relative abundance of bacteria within CF sputum samples was also observed in recent studies (Gast et al. 2011) (Zhao et al. 2012). This dominance occurred within the cohort and within individual samples. It was not observed for any other bacteria in the study and indicates the strong growth advantage of these major pathogens in the CF airways. In particular, relative abundances of more than 80 %, only observed for *P. aeruginosa* and *S. aureus*, are indicating the clinical relevance of these species. Besides these major pathogens, communities in sputum samples were further categorized by *Streptococcus* -1 and *R. mucilaginosa* which similarly not only exhibited high frequency of occurrence but also high relative abundances in individual samples. The OTU *Streptococcus -1* could not be affiliated to a single species but comparisons with SSCP fingerprints suggest that it may primarily represent *S. salivarius* and *S. parasanguinis* among other alpha-hemolytic *Streptococci*. Together with *R. mucilaginosa*, both species are known to be highly abundant in the healthy human oral microbiome (The Human Microbiome Project Consortium 2012). These species are commonly detected in CF sputum samples but their clinical relevance remains to be elucidated (Guss et al. 2011) (Gast et al. 2011). These four OTUs of the respective categories comprised the major proportion of most communities revealed in the cohort and characterize the associated samples in this regard. However, samples in the overlap between categories could not unequivocal be characterized and several outliers with distinct bacterial communities could not confidently associated with any category.

Elucidation of communities by polymicrobial consortia rather than single species is another approach with growing interest in the past years (Sibley et al. 2008) (Stressmann et al. 2012). Interactions between species in CF airways are well known and mixed biofilms have been studied especially between *P. aeruginosa* and *B. cepacia* (Eberl & Tümmler 2004). In respect to the complexity of communities found in the samples, different microbial subgroups were revealed by cluster analysis in the current study. Although a low significance level was applied, 14 distinct bacterial communities were determined in 72 samples which indicated the complexity of respiratory tract communities in CF patients. Previously observed outliers with more distinct bacterial communities and not well ordinated within the four previous categories, were comprised within these subgroups or defined their own group. Examples of such rather unique communities with great distance to other samples are the one with high abundant *Nocardia sp.* or *B. petrii*/*G. haemolysin*. The impact of individual species beyond

the well known pathogens can be elucidated much more detailed by the determination of such microbial groups. To further elucidate the persistence of polymicrobial consortia, sputum samples from a subcohort of patients were collected repeatedly and revealed certain unfrequently detected communities to be stable over a long period of time. In particular, the persistence in high proportions of distinct consortia including the strict anaerobic bacteria *S. millerii* and *F. nucleatum* was demonstrated and their pathogenic potential will be discussed below.

2.5.3 Clinical relevance of bacterial species is depending on consortia

Cystic fibrosis used to be a lung disease of young children but more recently has become a complex disease in adults (Davies et al. 2007). Improved diagnostics and treatment increased life expectancy and quality of patients but causative agents of respiratory infections are also known to change with age (Cystic Fibrosis Foundation Patient Registry 2012). Certainly, *P. aeruginosa* is the most important pathogen in CF and its pronounced adaptation to the airway of its host has been subject of many studies (Rau et al. 2012) (Folkesson et al. 2012). Together with *S. aureus* it is responsible for the majority of infections (LiPuma 2010). In contrast, *R. mucilaginosa* and *S. salivarius*/*S. parasanguinis* which were commonly detected together and were both not reported to be explicit causative agents of respiratory infections in CF. However, correlations between the microbial categories and clinical or host factors apparently fail to give unambiguous results for adult patients. Considering the pathogenicity of the respective species, patients with high abundant *S. aureus* and *P. aeruginosa* in their sputum were expected to show lower lung function values than patients from the third category. However, this was not generally the case. Furthermore, *P. aeruginosa* is known to be more prevalent in adults than in infant CF patients (Cystic Fibrosis Foundation Patient Registry 2012). A strong correlation of its abundance and the age of adult CF patients was not observed either. A recent study with pediatric and adult patients reported a strong relationship between the CFTR mutation and the microbial community (Cox et al. 2010). However, such a strong correlation was not observed for the cohort of adult patients. Neither of the factors like age, lung function or mutation allowed a prediction of the category of microbial community to be found within the sputum sample of the patient. These findings demonstrate the complexity of adult CF lung disease manifestations and illustrate that infections will not be understood by considering single pathogens only. Respiratory communities in CF are influenced by a number of physiological or environmental factors and suggest a more comprehensive view on the respiratory infections for adult patients.

Such a more comprehensive view was introduced by Rogers et al. (Rogers et al. 2009), who discussed the concept of bacterial communities as one pathogenic entity in the context of CF. The idea of pathogenic bacterial communities to cause infections rather than single species was introduced previously (Jenkinson & Lamont 2005). The current study, albeit small, demonstrates bacterial consortia from which some may act as pathogenic entities. For each subgroup, characteristic species were determined. Some communities were strongly dominated (relative abundance ≥ 50 %) by bacterial species with an unclear pathogenic potential as single infectious agents in CF, like *Nocardia* -1, *S. millerii* and *Fusobacterium nucleatum* (Bolstad et al. 1996) (LiPuma 2010) (Hauser et al. 2011). These species were infrequently found in the cohort suggesting specific conditions promoting their growth in individual communities. Particularly, species from the anaerobic *S. millerii* group are thought to be underestimated infectious agents and observed co-colonization with *P. aeruginosa* leading to speculation of polymicrobial interactions resulting in enhanced virulence in CF (Parkins et al. 2008). Communities with highly abundant *S. millerii* were observed to be persistent and the defined respective microbial subgroup accordingly comprised both bacteria as characteristic species. Also, communities characterized by high abundances of the strict anaerobic *F. nucleatum* (subgroup F) were associated with decreased lung functions. Sequent sputum samples of one patient revealed a stable community over eight months and suggesting persistence of *F. nucleatum* with high abundances in the communities. Although, it is regularly isolated from healthy oral body sites, this species is known to be involved in serious infections with a strong ability to form co-aggregates with other species (Bolstad et al. 1996). Strongly decreased lung functions were also associated with the unique community dominated by *B. petrii* and *G. haemolysins*. Regarding their clinical relevance, *B. petrii* was found to be persistent in one patient with a chronic pulmonary disease but previously described to lack pathogenic potential and *G. haemolysins*, known to be part of the normal oral flora but with pathogenic potential, was associated with CF before (Blitman et al. 2007) (Bittar et al. 2008) (Gross et al. 2008) (Le Coustumier et al. 2011).

Overall, distinct microbial communities were observed in the CF cohort, from which some polymicrobial entities may have pathogenic potential. Besides the limited number of species considered to be major CF pathogens, more recent studies revealed a wide range of bacteria in clinical samples from CF patients (Bittar & Rolain 2010) (Stressmann et al. 2011). Correlations between certain microbial subgroups and strongly decreased lung functions were found suggesting that these bacterial consortia have an influence on patient health and results

of the current study support the idea of enhanced pathogenicity due to interactions between species. Likewise, these findings indicate under-diagnosed species, like the strict anaerobic bacteria *S. millerii* and *F. nucleatum*, to be of clinical relevance and associated with infection manifestation and decline of health in CF patients.

2.6 References

Amélia Camarinha Silva (2011) *Dissertation:* Bacterial community structure and temporal dynamics of the human anterior nares - Chapter IV. Technischen Universität Carolo‐Wilhelmina, Braunschweig.

Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G & Neufeld JD (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* 77: 3846–3852.

Bassam BJ, Caetano-Anollés G & Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80–83.

Bittar F, Richet H, Dubus J-C, Reynaud-Gaubert M, Stremler N, Sarles J, Raoult D & Rolain J-M (2008) Molecular Detection of Multiple Emerging Pathogens in Sputa from Cystic Fibrosis Patients. *PLoS ONE* 3: e2908.

Bittar F & Rolain J-M (2010) Detection and accurate identification of new or emerging bacteria in cystic fibrosis patients. *Clin. Microbiol. Infect.* 16: 809–820.

Blitman NM, Steiner AM, Bell MD & Wilks-Gallo L (2007) Pulmonary nodules associated with Gemella bacteremia: CT findings in two children with osteosarcoma. *J Thorac Imaging* 22: 182–184.

Bolstad AI, Jensen HB & Bakken V (1996) Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. *Clin Microbiol Rev* 9: 55–71.

Boucher RC (2004) New concepts of the pathogenesis of cystic fibrosis lung disease. *Eur. Respir. J.* 23: 146–158.

Bray JR & Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* 27: 325–349.

Caporaso JG et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6: 1621–1624.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N & Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108 Suppl 1: 4516–4522.

Cleary DFR, Smalla K, Mendonça-Hagler LCS & Gomes NCM (2012) Assessment of variation in bacterial composition among microhabitats in a mangrove environment using DGGE fingerprints and barcoded pyrosequencing. *PLoS ONE* 7: e29380.

Cole JR et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37: D141–145.

Le Coustumier A, Njamkepo E, Cattoir V, Guillot S & Guiso N (2011) Bordetella petrii Infection with Long-lasting Persistence in Human. *Emerg Infect Dis* 17: 612–618.

Cox MJ et al. (2010) Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS ONE* 5: e11044.

Cystic Fibrosis Foundation Patient Registry (2012) *2011 Annual Data Report*. Cystic Fibrosis Foundation, Bethesda, Maryland http://www.cff.org/UploadedFiles/research/ClinicalResearch/2011-Patient-Registry.pdf (Accessed January 8, 2013).

Davies JC, Alton EWFW & Bush A (2007) Cystic fibrosis. *BMJ* 335: 1255–1259.

Eberl L & Tümmler B (2004) Pseudomonas aeruginosa and Burkholderia cepacia in cystic fibrosis: genome evolution, interactions and adaptation. *Int. J. Med. Microbiol.* 294: 123–131.

Eichler S, Christen R, Höltje C, Westphal P, Bötel J, Brettar I, Mehling A & Höfle MG (2006) Composition and Dynamics of Bacterial Communities of a Drinking Water Supply System as Assessed by RNA- and DNA-Based 16S rRNA Gene Fingerprinting. *Appl Environ Microbiol* 72: 1858–1872.

Flanagan JL et al. (2007) Loss of Bacterial Diversity during Antibiotic Treatment of Intubated Patients Colonized with Pseudomonas aeruginosa. *J. Clin. Microbiol.* 45: 1954–1962.

Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Høiby N & Molin S (2012) Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. *Nat. Rev. Microbiol.* 10: 841–851.

Gast CJ van der, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J & Bruce KD (2011) Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME Journal* 5: 780–791.

Ghozzi R et al. (1999) Capillary Electrophoresis–Single-Strand Conformation Polymorphism Analysis for Rapid Identification of Pseudomonas aeruginosa and Other Gram-Negative Nonfermenting Bacilli Recovered from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 37: 3374–3379.

Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL & Singh PK (2012) Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 109: 13769–13774.

Gross R et al. (2008) The missing link: Bordetella petrii is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. *BMC Genomics* 9: 449.

Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S & Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 5: 20–29.

Haanperä M, Jalava J, Huovinen P, Meurman O & Rantakokko-Jalava K (2007) Identification of Alpha-Hemolytic Streptococci by Pyrosequencing the 16S rRNA Gene and by Use of VITEK 2. *J. Clin. Microbiol.* 45: 762–770.

Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5: 235–237.

Harrison F (2007) Microbial ecology of the cystic fibrosis lung. *Microbiology (Reading, Engl.)* 153: 917–923.

Hauser AR, Jain M, Bar-Meir M & McColley SA (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clin Microbiol Rev* 24: 29–70.

Hogardt M (2006) *MIQ 24: Atemwegsinfektionen bei Mukoviszidose: Qualitätsstandards in der mikrobiologisch-infektiologischen Diagnostik*. Urban & Fischer by Elsevier.

Hori T, Haruta S, Ueno Y, Ishii M & Igarashi Y (2006) Direct comparison of single-strand conformation polymorphism (SSCP) and denaturing gradient gel electrophoresis (DGGE) to characterize a microbial community on the basis of 16S rRNA gene fragments. *J. Microbiol. Methods* 66: 165–169.

Ifp Institut für Produktqualität (2008) *PCRFast® Mycobacterium tuberculosis for Human Diagnostics*. Berlin, Germany http://novatec-id.com/pdf/pcr/englisch/PCRFast%20Mycobacterium%20tuberculosis_080924.pdf (Accessed January 8, 2013).

Jenkinson HF & Lamont RJ (2005) Oral microbial communities in sickness and in health. *Trends Microbiol.* 13: 589–595.

Klepac-Ceraj V et al. (2010) Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environmental Microbiology* 12: 1293–1303.

LiPuma JJ (2010) The Changing Microbial Epidemiology in Cystic Fibrosis. *Clin. Microbiol. Rev.* 23: 299–323.

Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW & Boucher RC (1998) Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 95: 1005–1015.

Meyer M & Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010: pdb.prot5448.

Parkins MD, Sibley CD, Surette MG & Rabin HR (2008) The Streptococcus milleri group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatr. Pulmonol.* 43: 490–497.

Pedrós-Alió C (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* 4: 449–466.

Rau MH, Marvig RL, Ehrlich GD, Molin S & Jelsbak L (2012) Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. *Environ. Microbiol.* 14: 2200–2211.

Rogers GB, Carroll MP & Bruce KD (2009) Studying bacterial infections through culture-independent approaches. *J Med Microbiol* 58: 1401–1418.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G & Bruce KD (2004) Characterization of Bacterial Community Diversity in Cystic Fibrosis Lung Infections by Use of 16S Ribosomal DNA Terminal Restriction Fragment Length Polymorphism Profiling. *J. Clin. Microbiol.* 42: 5176–5183.

Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC & Surette MG (2008) A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci U S A* 105: 15070–15075.

Soergel DAW, Dey N, Knight R & Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* 6: 1440–1444.

Stewart, Camper, Handran, Huang & Warnecke (1997) Spatial Distribution and Coexistence of Klebsiella pneumoniae and Pseudomonas aeruginosa in Biofilms. *Microb. Ecol.* 33: 2–10.

Stressmann FA et al. (2011) Analysis of the bacterial communities present in lungs of patients with cystic fibrosis from American and British centers. *J. Clin. Microbiol.* 49: 281–291.

Stressmann FA et al. (2012) Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* 67: 867–873.

The European Respiratory Society (1993) Standardized lung function testing. Official statement of the European Respiratory Society. *Eur Respir J Suppl* 16: 1–100.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207–214.

Tunney MM et al. (2008) Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* 177: 995–1001.

VanDevanter DR & LiPuma JJ (2012) Microbial diversity in the cystic fibrosis airways: where is thy sting? *Future Microbiol* 7: 801–803.

Worlitzsch D, Rintelen C, Böhm K, Wollschläger B, Merkel N, Borneff-Lipp M & Döring G (2009) Antibiotic-resistant obligate anaerobes during exacerbations of cystic fibrosis patients. *Clin. Microbiol. Infect.* 15: 454–460.

Wos-Oxley ML, Plumeier I, Eiff C von, Taudien S, Platzer M, Vilchez-Vargas R, Becker K & Pieper DH (2010) A poke into the diversity and associations within human anterior nare microbial communities. *The ISME Journal* 4: 839–851.

Zhao J et al. (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci. U.S.A.* 109: 5809–5814.

2.7 Supplementary Material

forward primers	sequence (5'->3')
V3_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AAGCCT**CAATTACCGCGGCTGCTGG
V3_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AAGTTA**CAATTACCGCGGCTGCTGG
V3_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AATACG**CAATTACCGCGGCTGCTGG
V3_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AATCGA**CAATTACCGCGGCTGCTGG
V3_5	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AATGAT**CAATTACCGCGGCTGCTGG
V3_6	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACCAAT**CAATTACCGCGGCTGCTGG
V3_7	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACCGTC**CAATTACCGCGGCTGCTGG
V3_8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACCTCG**CAATTACCGCGGCTGCTGG
V3_9	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACGAGG**CAATTACCGCGGCTGCTGG
V3_10	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACTATA**CAATTACCGCGGCTGCTGG
V3_11	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**ACTTAC**CAATTACCGCGGCTGCTGG
reverse primer	
IlluRevAdap_V3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTACGGGAGGCAGCAG

Table S1: Primers used for library preparation in Illumina-based sequencing. Underlined letters denote complementary sequences for the variable region V3 of the 16S rRNA gene. Eleven different barcodes were used, each indicated with bold letters within the primer sequence. In a second PCR, amplicons for libraries were accomplished by adding Illumina-specific indices.

Table S2:

OTU Illumina	Identity in %	closest representative	species by SSCP	pos. samples (total no.)	min. abund. in %	max. abund. in %
Streptococcus -1	100	KC632197.1 Streptococcus parasanguinis		63	0.86	51.67
Rothia mucilaginosa	100	KC632201.1Rothia mucilaginosa	99% R. mucilaginosa DQ409140.1	60	0.54	50.06
Pseudomonas -1	100	HF572851.1 Pseudomonas aeruginosa	100% P. aeruginosa CP003149.1	57	0.63	93.31
Rothia -2	100	KC632226.1 Rothia dentocariosa		36	0.72	25.77
Granulicatella -1	100	JF803551.1 Granulicatella para-adiacens		33	0.53	5.87
Prevotella -1	100	JN867317.1 Prevotella melaninogenica		30	0.55	38.32
Staphylococcus -1*	100	KC465401.1 Staphylococcus aureus	99% S. aureus AB680391.1	27	0.60	87.81
Gemella haemolysans	100	HE974940.1 Gemella haemolysans	100% G.haemolysans NR025903.1	22	0.57	31.49
Gemella -2	100	AY757357.1 Gemella sanguinis		22	0.51	4.28
Atopobium -1	100	GU425926.1 Atopobium rimae		18	0.55	3.23
Streptococcus -3	100	KC632205.1 Streptococcus sanguinis		18	0.54	8.77
Streptococcus millerii	100	KC632214.1 Streptococcus anginosus		17	0.54	75.53

Streptococcus -2	99	KC632193.1 Streptococcus salivarius		16	0.60	10.10
Actinomyces graevenitzii	97	AY866429.1 Actinomyces graevenitzii		16	0.58	14.51
Actinomyces -1	100	HQ850579.1 Actinomyces turicensis		15	0.57	14.16
Actinomyces -2	97/100	HQ616392.1 Actinomyces sp. / JQ031125.1 Actinomyces funkei		13	0.54	5.53
Veillonella atypica	100	KC632238.1 Veillonella dispar		13	0.52	13.83
Veillonella -2	100	HM596287.1 Veillonella sp.		11	0.70	5.98
Fusobacterium	100	NR_074412.1 Fusobacterium nucleatum	98% F.nucleatum AB588016.1	11	0.55	53.52
Achromobacter -1	100	KC010531.1 Achromobacter xylosoxidans	99% A. xylosoxidans EU266588.1	6	0.77	46.35
Stenotrophomonas maltophilia	100	KC683776.1 Stenotrophomonas maltophilia	99% S.maltophilia JX848739.1	6	0.62	15.41
Mogibacterium -1	100	KC632200.1 Mogibacterium diversum		5	0.51	1.31
Actinomyces -3	97/100	HQ616392.1 Actinomyces sp. / JQ031122.1 Actinomyces funkei		5	0.66	1.74
TM7	100	GQ422738.1 TM7 phylum sp. oral		5	0.52	1.93
Porphyromonas -1	100	GU409265.1 Porphyromonas sp.		4	0.76	1.58
Bifidobacterium longum	100	HE974928.1 Bifidobacterium longum		4	0.54	3.55
Stomatobaculum longum	100	HM120209.1 Stomatobaculum longum		4	0.51	4.36
Clostridiaceae -1	100	NR_074652.1 Clostridium phytofermentans		4	0.50	1.11
Prevotella pallens	100	KC632223.1 Prevotella pallens		4	0.55	0.90
Nocardia -1	100	KC478309.1 Nocardia farcinica		3	1.10	61.87
Solobacterium moorei	100	JX104033.1 Solobacterium sp.		3	0.57	0.70
Actinomyces -4	96/95	EF473992.1 Actinomyces sp. / JF803550.1 Actinomyces graevenitzii		3	3.36	13.04
Prevotella salivae	100	JN867237.1 Prevotella salivae		3	0.52	1.58
Granulicatella elegans	100	JN801174.1 Granulicatella elegans		3	0.60	0.81
Streptococcus mitis	100	GU326246.1 Streptococcus pneumoniae		3	0.74	1.17
Actinomyces -5	100	JX524818.1 Actinomyces oris		3	0.53	0.75
Lactobacillus -2	100	KC456368.1 Lactobacillus fermentum		3	0.60	1.40
Prevotella - 2	100	JN867292.1 Prevotella oris		2	5.64	12.27
Rothia-3	100	GQ900845.1 Rothia sp.		2	0.51	1.10
Haemophilus -2	100	JF506651.1 Haemophilus parainfluenzae		2	1.21	1.21
Streptoccocus -4	99	KC632199.1 Streptococcus australis		2	0.51	0.67
Veillonella -2	100	AB679112.1 Veillonella tobetsuensis		2	1.52	2.04
Megasphaera micronuciformis	98.59	JF803576.1 Megasphaera micronuciformis		2	0.50	0.83

Porphyromonas -2	100	JF803575.1 Porphyromonas sp.		2	1.05	2.49
Atopobium -2	100	KC297229.1 Atopobium parvulum		2	0.54	1.29
Prevotella tannerae	100	GU561345.1 Prevotella tannerae		2	0.52	1.26
Burkholderia -1	100	HF678361.1 Burkholderia cepacia	100% B. cepacia AB680641.1	1	8.60	8.60
Micrococcineae -1	100	KC346300.1 Streptomyces violaceorectus		1	1.12	1.12
Neisseria -2	100	KC178474.1 Neisseria mucosa		1	2.43	2.43
Acinetobacter -1	100	KC337241.1| Acinetobacter tjernbergiae		1	1.09	1.09
Neisseria flavescens	100	KC178511.1 Neisseria flavescens		1	1.61	1.61
Bordetella petrii	100	HE672085.1 Bordetella petrii		1	55.88	55.88
Prevotella -3	100	GU409603.1 Prevotella sp.		1	2.92	2.92
Veillonella -3	93	KC632234.1 Veillonella sp.		1	1.94	1.94
Scardovia wiggsiae	100	HM596282.1 Scardovia wiggsiae		1	0.54	0.54
Prevotella nanceiensis	100	JN867319.1 Prevotella nanceiensis		1	0.68	0.68
Capnocytophaga gingivalis	100	GU561334.1 Capnocytophaga gingivalis		1	1.81	1.81
Leptotrichia wadei	100	AB588021.1 Leptotrichia wadei		1	0.53	0.53
Lactobacillus kalixensis	100	FR683096.1 Lactobacillus kalixensis		1	1.39	1.39
Prevotella -4	98	GU413291.1 Prevotella sp.		1	2.28	2.28
Alloscardovia -1	100/97	AB425070.1 Alloscardovia sp. / AB437351.1 Bifidobacterium psychraerophilum		1	1.90	1.90
Propionibacterium acidifaciens	100	JF803561.1 Propionibacterium acidifaciens		1	0.60	0.60
Haemophilus influenzae	100	CP000057.2 Haemophilus influenzae		1	1.11	1.11
Leptotrichia -2	100	NR_074440.1 Leptotrichia buccalis		1	0.74	0.74
Propionibacterium acnes*	100	JQ435685.1 Propionibacterium acnes		1	3.59	3.59
			99% S. parasanguinis AY281078			
			100% S. salivarius FR873482.1			
			99% S. mitis FN568063.1			
			99% S. gordonii AY281088.1			
			100% S. sanguinis AB596946.1			
			99% S. peroris GU425263.1			
			99% S. oralis AF003932			
			99% S. anginosus AF306838.1			
			100% S. intermedius AF104671.1			

Table S2: OTUs observed with Illumina exhibiting a minimum relative abundance of 0.5 %. Sequence similarity revealed by BLAST is given as well as the closest representative (acquisition number and name) for each OTU. Comparison with SSCP fingerprinting allowed phylogenetic affiliation of some previously ambiguous OTUs to species level. Streptococcus species identified by SSCP are mentioned with sequence identity scores in percent as well as accession numbers of closest representatives in the end of the table. For Illumina-based data, number of positive samples for each OTU is given as well as minimum and maximum relative abundances in individual samples. *V1 instead of the V3 region was amplified in rare events for the OTUs Propionibacterium acnes as well as for Staphylococcus -1 in some samples. However, comparisons of the relative abundances assessed by repeatedly analyzed samples did not reveal any significant difference in the affected samples.

Figure S1 (below): Comparison of individual relative abundances for all samples assessed by NGS (^) and SSCP fingerprinting (°). Only OTUs with relative abundance of ≥ 5 % were considered for these comparisons. Each diagram shows 24 samples in total which means 12 distinct samples each one analyzed with both methods. Each colour represents one bacterial OTU according to the legend. OTU only defined in NGS are respectively labelled (^) as well as OTUs only defined in SSCP fingerprinting (°). OTUs unambiguously and equally assigned with both methods are accordingly marked with both labels (^°). Numbers indicate individual patients. Letters are according order of sputum collection.

Figure S1 (continuing):

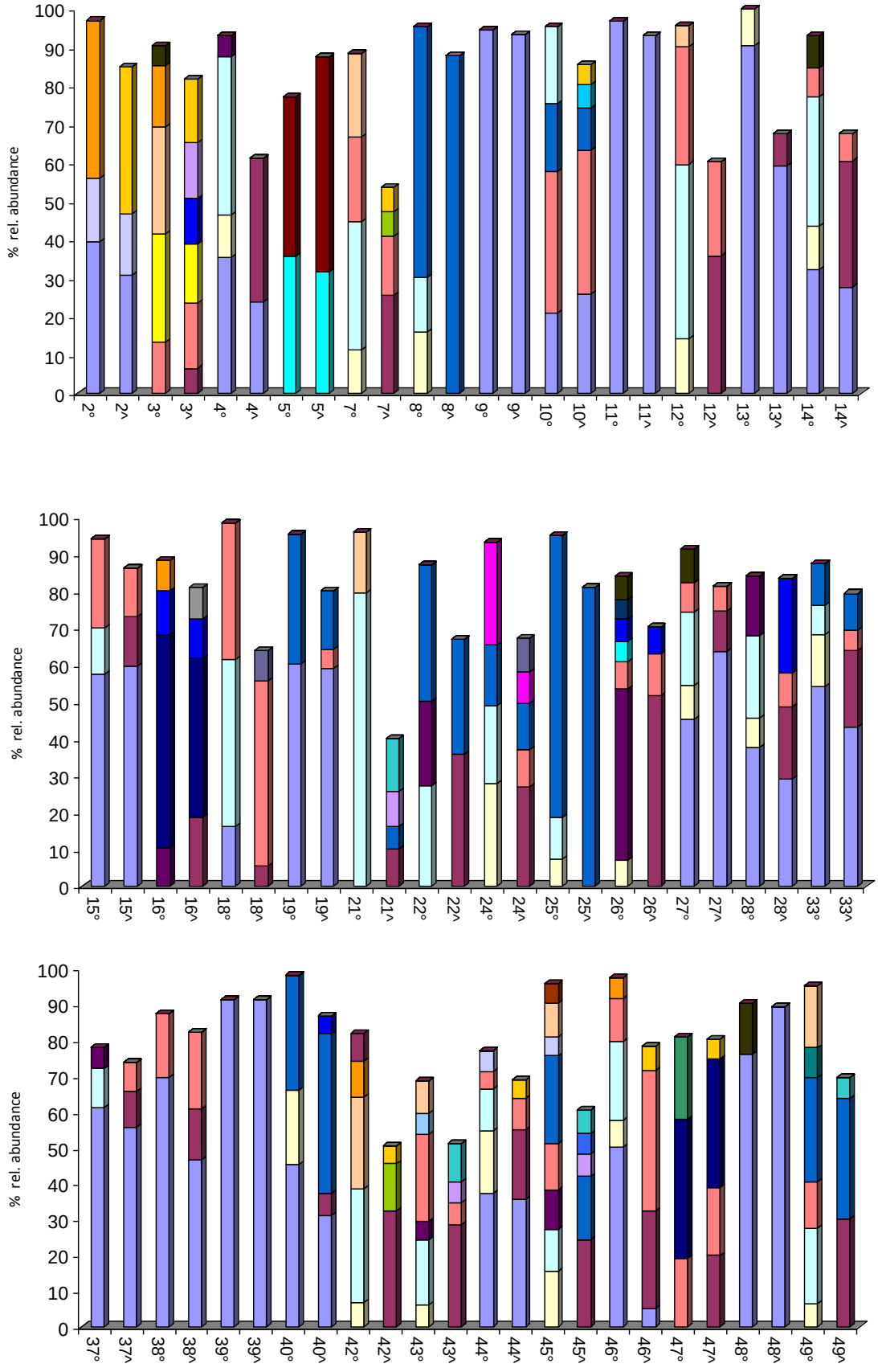


Figure S1 (continuing):

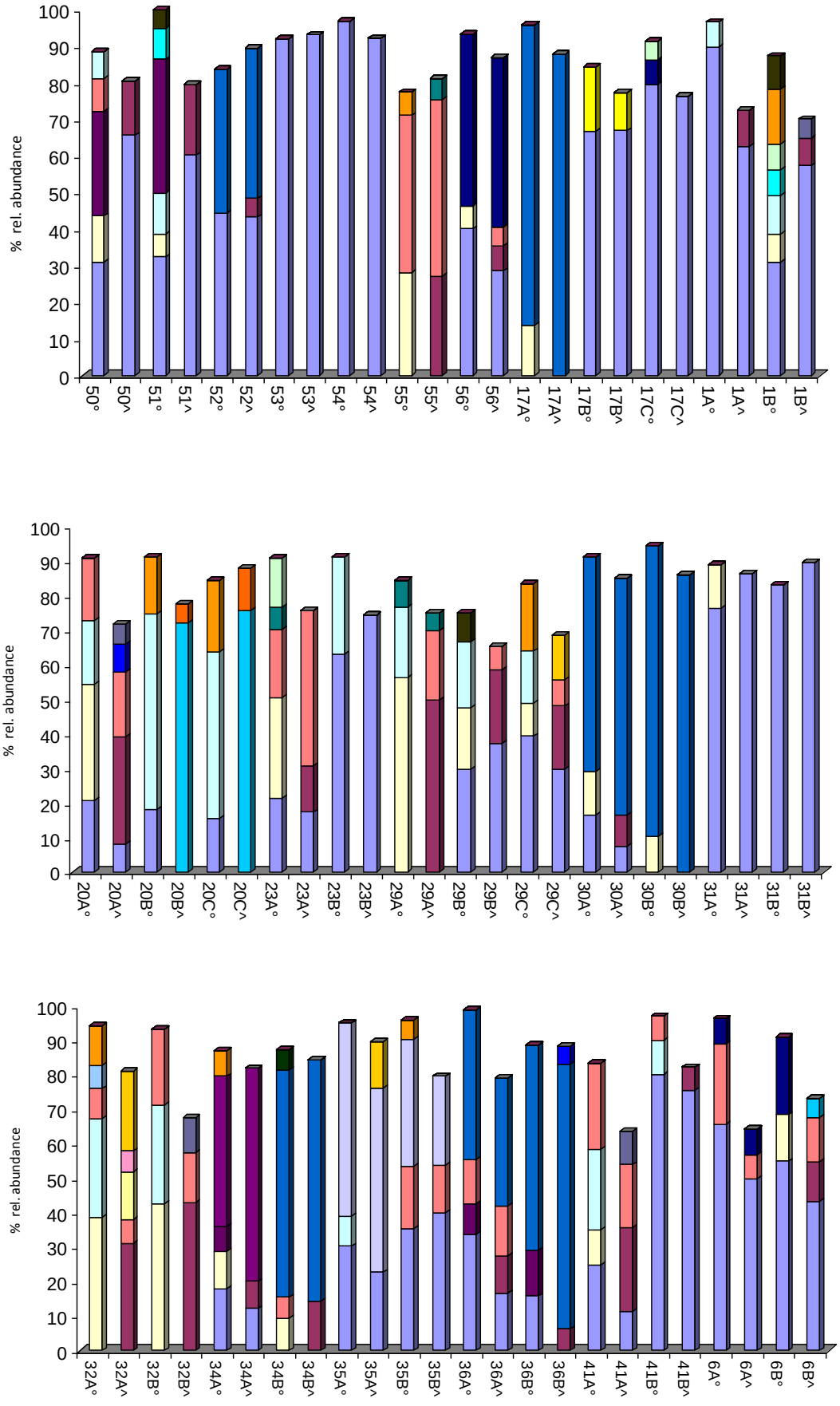


Figure S2: Histograms of RELATE analysis: Testing matched resemblance matrices based on separately analyzed compositional data sets. Similarity is measured by Spearman's Rho and data sets were permuted. Distribution of Rho values for permutations are shown in the histograms. The according frequencies of permutations are given on y-axes. Rho values of the actual sample are additionally shown in each histogram. For all analysis the number of permuted statistics greater than or equal to Rho value of the actual sample was zero.

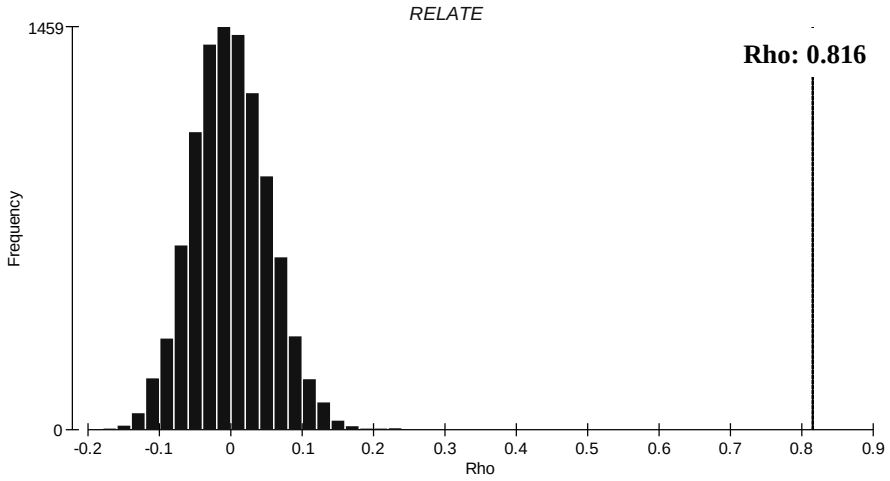


Figure S2a: Testing similarity of compositional data sets for 72 sputum samples assessed by NGS and by SSCP. Significance level of sample statistic: 0.01 %. Number of permutations: 9999.

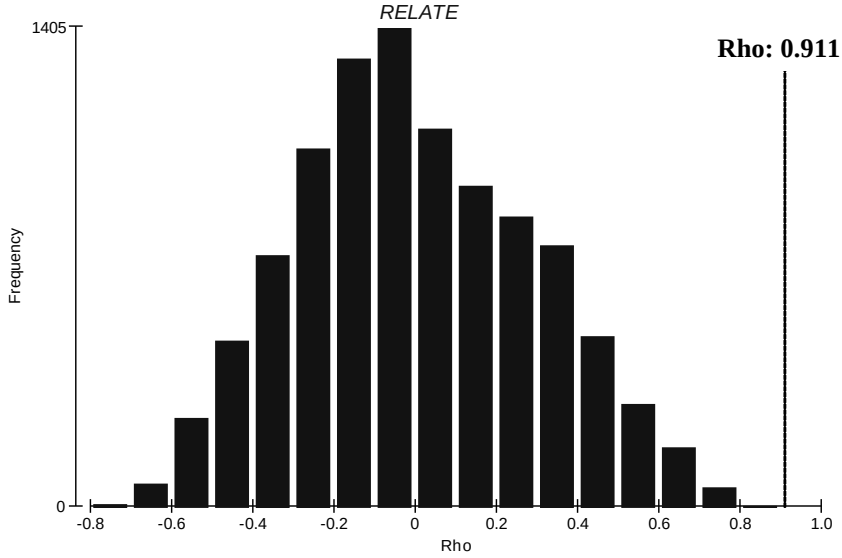


Figure S2b: Testing similarity of compositional data sets for different extractions of 8 sputum samples assessed by NGS. Significance level of sample statistic: 0.01 %. Number of permutations: 9999.

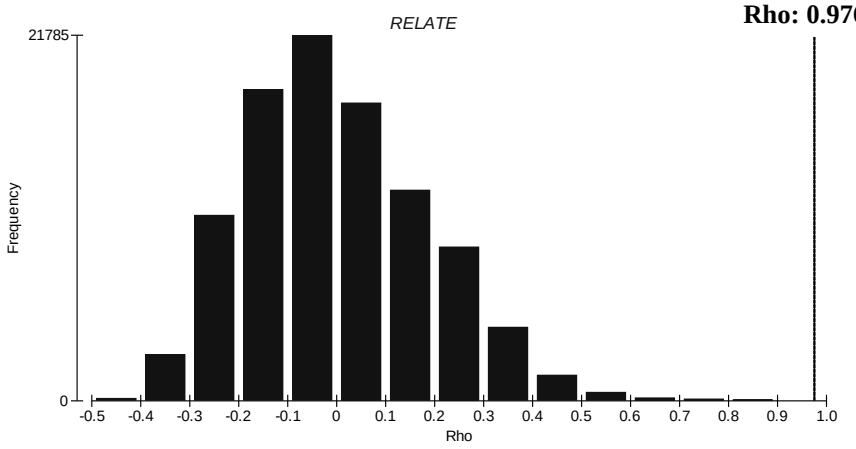


Figure S2c: Testing similarity of compositional data sets for different amplifications and preparations of 10 sputum samples assessed by NGS. Significance level of sample statistic: 0.001 %. Number of permutations: 99999.

Figure S4:

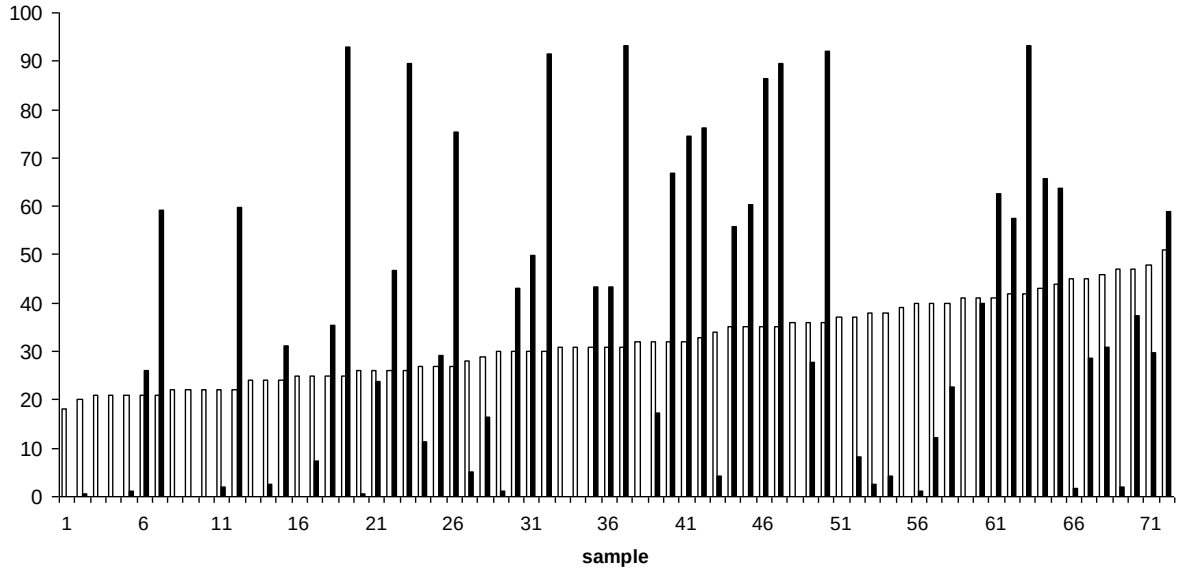
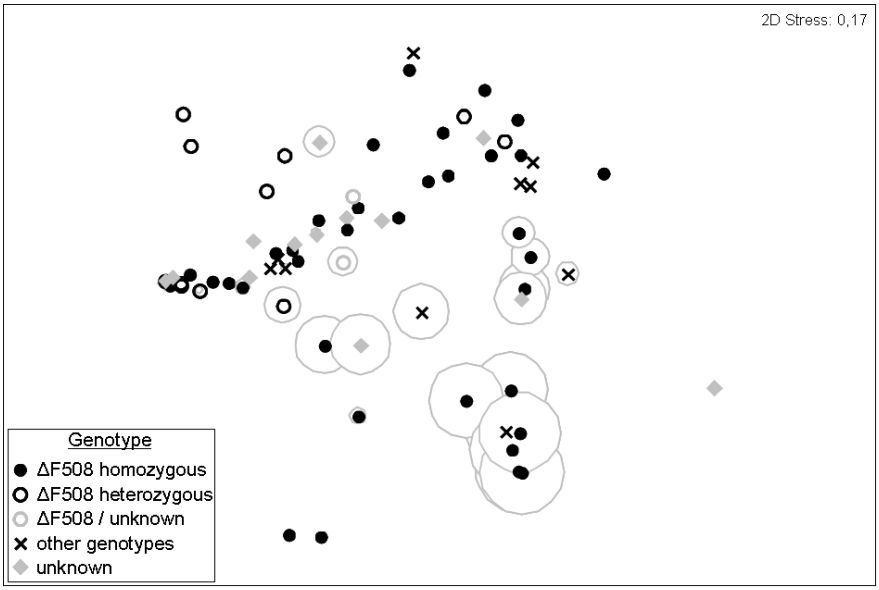


Figure S4: Comparison between age of patients (white bars) and relative abundance of *P. aeruginosa* (black bars). Samples are ordered by age of the patient at the individual time point of sputum collection. Age of patient is given in years and relative abundance of *P. aeruginosa* is given in % on y-axis. For each patient, both parameters are indicated. Missing bars for *P. aeruginosa* indicate its absence in the sample. A median relative abundance of 24.8 % was calculated for the bacteria. Median age at time point of sputum collection was 31 years.

Figure S5: CFTR mutations of patients superimposed on the previous MDS plot calculated for the community compositions of 72 sputum samples. Three classes of identified genotypes were determined for the cohort. Partly or completely unknown genotypes are accordingly labelled.



Chapter III

Characterization of fungal and bacterial communities in sputum samples from adult patients with cystic fibrosis

Rolf Kramer[1], Annette Sauer-Heilborn[2], Carlos A. Guzman[1],
Wolf-Rainer Abraham[1] and Manfred G. Höfle[1]

[1]Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2]Department of Pneumology, Hannover Medical School (MHH), Germany

3.1 Abstract

The current study presents major bacterial and fungal species associated with cystic fibrosis (CF). Molecular fingerprinting based on single-strand conformation polymorphism (SSCP) analysis revealed a rather different picture for bacteria and fungi detected in 72 sputum samples. Microorganisms were further taxonomically classified and the individual detection frequencies are presented. In contrast to bacteria, a large number of fungi were observed together with high fluctuation rates over time and between different patients, suggesting low colonization abilities of fungi in CF airways. Major bacterial pathogens were frequently observed, whereas fungal species with demonstrated pathogenicity in CF were detected at a low frequency. Overall, the current study provides for the first time an overview of the major bacterial and fungal species of an entire microbiome in a broader CF cohort.

3.2 Introduction

Cystic fibrosis (CF) is the most common inherited disease in the western human population and respiratory tract infections are the leading cause of death for these patients. Human airways are covered with a mucus layer that serves to trap inhaled objects and in a permanent clearance process the mucus is removed from the respiratory tract. In CF patients, disturbed ion transports in airway epithelia lead to a dysfunction of this clearance mechanism, resulting in conditions that promote microbial colonization (Matsui et al. 1998) (Boucher 2004). Improved treatments and diagnostic methods have led to a constantly increasing life expectancy during the past two decades and it is predicted that within the next years more than 50% of the population with CF will be adults (Cystic Fibrosis Foundation Patient Registry 2012). In turn, the occurrence and frequencies of pathogens are changing with age, resulting in new challenges in the healthcare of these patients. Major bacterial CF pathogens belong to the phylum *Proteobacteria* and include *Pseudomonas aeruginosa*, *Burkholderia cepacia* and *Stenotrophomonas maltophilia*. Together with *Staphylococcus aureus* from the phylum *Firmicutes* they are considered to be responsible for the majority of respiratory tract infections in adult CF patients (LiPuma 2010) (Hauser et al. 2011). Besides bacteria, fungi also exhibit the potential to cause lung diseases, like the allergic bronchopulmonary aspergillosis (ABPA). Already in the 1970s, culturable fungi were observed in over 80% of sputum and human lung samples which were taken randomly from necropsies (reason of death not specified) (Mullins & Seaton 1978). Despite the high exposure, the number of reported fungal lung infections is remarkably low but the real prevalence and role of fungal pathogens in CF patients still remains to be elucidated (Pihet et al. 2009).

Additionally, the clinical relevance of several bacterial species associated with CF sputum is yet to be defined. Culture-independent methods revealed a broad spectrum of bacteria that are frequently detected in sputum and a polymicrobial concept for the development of respiratory diseases in the context of CF was suggested (Rogers et al. 2009). In such a polymicrobial concept, the pathogenicity of certain species depends on the interaction with others and, as a consequence, the community as a whole has to be considered pathogenic rather than single species (Jenkinson & Lamont 2005) (Delhaes et al. 2012). Ecological and physiological factors related to one microorganism may promote the growth and colonization ability of others. In this regard, fungal species also present in the human respiratory tract have to be considered in microbiological studies. In contrast to the bacterial community, a comprehensive view of fungal species being part of the human respiratory microflora was hardly approached. Only in a recent study a fungal microbiome of healthy individuals was

proposed for the oral cavity (Ghannoum et al. 2010). Interactions between fungi and bacteria have been reported and shown in more detail for the yeast *Candida albicans* and *P. aeruginosa* (Cugini et al. 2007) (McAlester et al. 2008) (De Sordi & Mühlschlegel 2009). A negative correlation between the *Candida spp.* load and the bacterial species diversity of the salivary microbiome in elderly healthy individuals was observed (Kraneveld et al. 2012). In the context of cystic fibrosis, fungal and bacterial species out of the same sputum samples were recently analyzed from four patients and correlations drawn with the clinical data (Delhaes et al. 2012). Overall, knowledge about fungal communities associated with the human respiratory tract is poor. Occurrence as well as exposure rates for the fungi first have to be determined before conclusions about their impact on respiratory microbial communities or on the health of infected patients can be made.

The aim of the current study was to provide an overview about the fungal and bacterial taxa present in CF sputum using a culture-independent approach. This taxonomic survey was further compared with clinical abnormalities observed in the patients. In total, 72 sputum samples from 56 adult CF patients were analyzed. Species richness from the CF cohort was estimated and detection frequencies given for all fungal and bacterial taxa which were further taxonomically classified. The current study is the first to present the major taxa of an entire microbiome in a broader CF cohort, including fungi and bacteria.

3.3 Material and Methods

3.3.1 Sample collection

Sputum samples were collected in sterile containers from 56 CF patients recruited in the cystic fibrosis ambulance centre of the Hannover Medical School (MHH; Hannover, Germany), including a subgroup of 13 patients who provided sputum twice (n=10) or three times (n=3) within the sampling period of 2 years. In total, 72 separate sputum samples were collected. Ethical approval for the current study was granted by the local health authority ethics committee. Sputum collection was done during the routine medical examination. All patients were between 18-51 years old and a median age of 31 was calculated for the cohort. Both genders were equally represented (48.2% female and 51.8% male). Signs of acute infections were observed for some patients, none of them were in a critical health condition at the time of sputum collection. Containers with sputum were stored at -20°C before DNA extraction.

3.3.2 Sputum preparation and DNA extraction

An optimized protocol for the extraction of fungal and bacterial DNA was developed by modifying the manufacturer's instructions of the kit and by adaptation of guidelines for the preparation of sputum samples recommended for detection of *Mycobacterium tuberculosis* in sputum (ifp Institut für Produktqualität 2008). Briefly, sputum samples were aliquoted and boiled for 15 minutes. To decrease the viscosity, a cysteine buffer (2% NaOH, 1.45% sodium citrate and 0.5% N-acetylcysteine) was added in the same volume and the solution was mixed for 40 minutes. Milli-Q water was added to a final volume of 15 ml and centrifuged for 30 minutes at 4000 g. The supernatant was discarded and the pellet was resuspended in 300 µl of lysis buffer (20 mM Tris-Cl with pH 8.0, 2 mM sodium EDTA, 1.2% Triton® X-100). 6 mg of lysozyme (SERVA; Heidelberg, Germany) was added and incubated for 30 min at 37°C. To the total volume of the solution 0.5% of β-mercaptoethanol was given together with 50 U of lyticase (Sigma-Aldrich; Germany). The solution was incubated for another 45 min at 37°C before centrifugation at 12.000 rpm for 10 min. The pellet was resuspended in 300 µl Lyse T buffer from GeneMATRIX Tissue & Bacteria DNA purification kit (EurX Roboklon; Berlin, Germany) and 20 µl of proteinase K (Qiagen; Hilden, Germany) were added. After incubation of 2 hours at 56°C, the DNA was extracted by following the instructions of the manufactures. DNA extracts were kept frozen at -20°C for further analysis.

3.3.3 PCR amplification and preparation of ssDNA

PCR amplification of parts of the rRNA gene from bacteria and fungi was performed. For bacteria, the primers 27F (5´-AGAGTTTGATCMTGGCTCAG-3') and 521R (5'-ACCGTGGCTGCTGGCAC-3') were used to amplify parts of the 16S rRNA gene, including the variable regions V1-V3 of the small subunit of the ribosome (SSU) with an amplicon size between 459 bp for *Nocardia sp.* and 505 bp for *Veillonella sp.* (Wos-Oxley et al. 2010). PCR was carried out using 50 ng DNA of sputum extractions in a final volume of 50 µl, starting with an initial denaturation for 15 min at 95°C. A total of 30 cycles (1 min at 95°C, 40 sec at 56°C, and 1 min at 72°C) was followed by a final elongation for 10 min at 72°C. For fungi, a semi-nested PCR strategy was chosen, using the forward primers ITS1 (5´-TCCGTAGGTGAACCTTGCGG-3'), ITS4 (5´-TCCTCCGCTTATTGATATGC-3´) and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3') for reverse primers respectively (Schoch et al. 2012). High length heterogeneity for the internal transcript spacer (ITS) regions are known, amplicon sizes in the current study varied from 134 bp for *Yarrowia lipolytica* to 475 bp for *Candida glabrata* (Chen et al. 2001). The first PCR was carried out using 50 ng DNA from sputum extracts in a final volume of 50 µl. By adding 3µl of $MgCl_2$ (25 mM), PCR efficiency was substantially improved. Starting with an initial denaturation for 15 min at 95°C, 20 cycles (1 min at 95°C, 1 min at 58°C, and 1 min at 72°C) were followed by a final elongation for 10 min at 72°C. Amplification of the final target region, ITS between the 18S and 5.8S, was achieved by using 1 µl of the previous PCR in a final volume of 50 µl. Initial denaturation for 15 min at 95°C was followed by a total of 30 cycles (30 sec at 95°C, 20 sec at 62°C, and 30 sec at 72°C) and a final elongation for 10 min at 72°C. 1.5 U of HotStarTaq DNA polymerase was used for all amplifications (Qiagen). For single-strand DNA (ssDNA) preparation, reverse primer 521R and ITS2 were 5´-biotin labeled and magnetic streptavidin coated beads (Promega, Madison, Wis.) were applied to obtain ssDNA from the PCR amplicons according to Eichler et al. (Eichler et al. 2006).

3.3.4 SSCP fingerprints and sequencing of individual ssDNA bands

Dried pellets of ssDNA were resuspended in 7 µl of gel loading buffer (95% formamide, 10 mM NaOH, 0.25% bromphenol blue, 0.25% xylene cyanol). After incubation for 10 min at 95°C, the ssDNA samples were stored on ice and loaded onto a non-denaturing polyacrylamide-like gel (0.6x MDE gel solution; Cambrex BioScience, Rockland, Maine) for SSCP electrophoresis. SSCP fingerprints of bacterial DNA were obtained at 20°C and 400 Volt for 21 hours on 20 cm glass plates and at 20°C and 700 Volt for 20 hours for fungal

DNA on 55 cm glass plates respectively. More details on the SSCP fingerprints are given in Eichler et al. (Eichler et al. 2006). The gel was silver stained according to the method described by Bassam et al. (Bassam et al. 1991). SSCP fingerprints were digitized using a HP Scanjet G4050 scanner. Bands with an intensity ≥1% of the total lane were considered for further analysis and excised from the SSCP gel, boiled in Tris buffer (10 mM Tris-HCl, 5 mM $MgCl_2$, 5 mM KCl, 0.1% Triton X-100, pH 9.0) and re-amplified according to the conditions described above. Purified amplicons were sequenced by cycle sequencing (ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit; Applied Biosystems, Foster City, Calif.), products were purified again with the BigDye Terminator purification kit (Qiagen) and analyzed using capillary electrophoresis with fluorescence detection (ABI Prism 3100 Genetic Analyzer).

3.3.5 Data processing and phylogenetic analysis

Phylogenetic analysis of sequences was performed with the NCBI Tool BLAST/blastn and Ribosomal Data Base Project (RDP) Seqmatch Tool (only for bacteria) (Cole et al. 2009). The closest taxonomic groups were determined by sequence similarity and height of ssDNA bands in the molecular fingerprints. For fungi, species identification was defined for sequences with a similarity of at least 97%. Operational taxonomic units (OTUs) with ambiguous BLAST results were defined according to the closest shared taxonomic level. Taxonomic classification of fungi followed the outline published by Hibbett at al. (Hibbett et al. 2007). Classification of fungi is originally based on morphological features and developed further by genetic and physiological analysis. Therefore, molecular analysis of parts of the ribosomal RNA operon does allow identification of fungi but associated phylogenetic trees may not always display closely related species in the same clade. For bacteria, species identification was defined for sequences with a similarity of >97% and ssDNA bands of the same heights were considered to be the same species. Further, phylogenetic distance analysis was performed to confirm OTU determination (data not shown). Phylogenetic distances of sequences were calculated with Jukes-Cantor algorithm and displayed in Neighbour-joining trees using MEGA version 5 (Tamura et al. 2011). Species accumulation curves were calculated using AccuCurve 1.0 (Drozd & Novotny 2010). Incidence-based datasets were applied, permutated 250 times and the resultant curves were averaged.

3.4 Results

3.4.1 Comparison of fungal and bacterial species richness in CF cohort

The incidence of bacterial and fungal species was assessed in 72 separate sputum samples, collected over a period of 2 years. Molecular fingerprints of the microbial communities from sputum were generated using SSCP electrophoresis, followed by sequencing of all major bands (**Figure 1**).

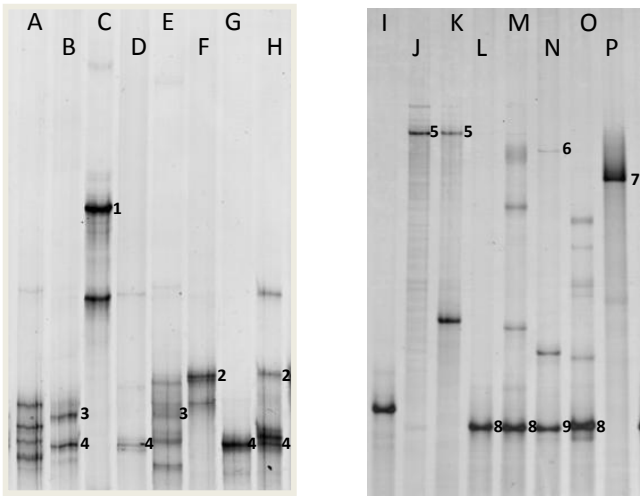


Figure 1: Molecular fingerprints by SSCP electrophoresis from bacteria (left) and fungi (right). Each lane (A-H; I-P) represents one sputum sample. Each band represents ssDNA from one microbial species. Bands at the same height were the same species. Numbers indicate some species identified by sequencing of the respective bands: For bacteria 1= *Bordetella petrii*, 2= *Staphylococcus aureus*, 3= *Streptococcus salivarius*, 4= *Pseudomonas aeruginosa*; for fungi 5= *Saccharomyces cerevisiae*, 6= *Candida glabrata*, 7= *Exophiala dermatitidis*, 8= *Candida albicans* and 9= *Candida dubliniensis*. Discrimination between *C. albicans* and *C. dubliniensis* was only achieved by sequencing.

The resultant sequences were aligned and the OTUs defined. In total, 60 fungi and 38 bacteria were considered as OTUs and species affiliation identified by comparative sequence analysis (**Table S1, S2)**. Individual species counts per sample showed a median of 4 bacterial and 2 fungal OTUs per sputum sample (**Figure 2**). In most of the 72 sputum samples between 3-6 bacterial species and 1-3 fungal species were observed. For bacteria and fungi, a maximum number of 11 species in one sample was detected. At least one bacterial OTU was detected in each sample, whereas few sputum samples exhibit no fungal species count.

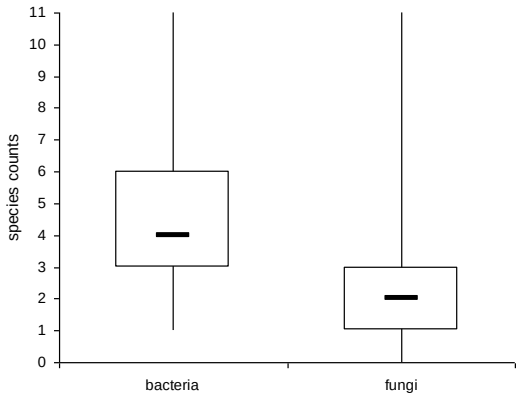


Figure 2: Boxplots of species counts per sample for bacteria and fungi. Bars within boxes indicate the median number of OTUs. End of whiskers represent the minimum and maximum of species counts found in one sample.

In order to estimate richness for both types of microbes in the cohort, species accumulation curves were calculated using observed species counts (S) and counts for species only occurring in one sample (Q1; unique species) (**Figure 3**). Comparison of both curves for S showed a rather different picture for fungi and bacteria. For bacteria, the curve strongly rises in the beginning, indicating a high number of bacterial species in a relatively small number of samples. With increasing sample number the curve flattens and indicates in the end only a small incremental increase in species number if more samples would be analyzed. For fungi, no flattening of the curve was observed. A small number of samples showed a low incidence of species but a continuous rise of the curve indicates a steady increase in fungal OTUs. Considering that in a comprehensive sampling the number of unique species should be as low as possible, the trends for the curves of Q1 are more informative for the estimation of species richness. For bacteria, a flat curve similarly shows only a small increase of unique species when taking more samples into account and indicates in the end no substantial increase of new unique species with further sampling. In 72 sputum samples, 15 bacterial species were only occurring in one sample (unique species). Again, a rather different picture was observed for fungi: the curve for Q1 indicates a continuous rise of unique species with increasing sample numbers. After 72 analyzed sputum samples, 42 out of 60 fungal OTUs were occurring only once.

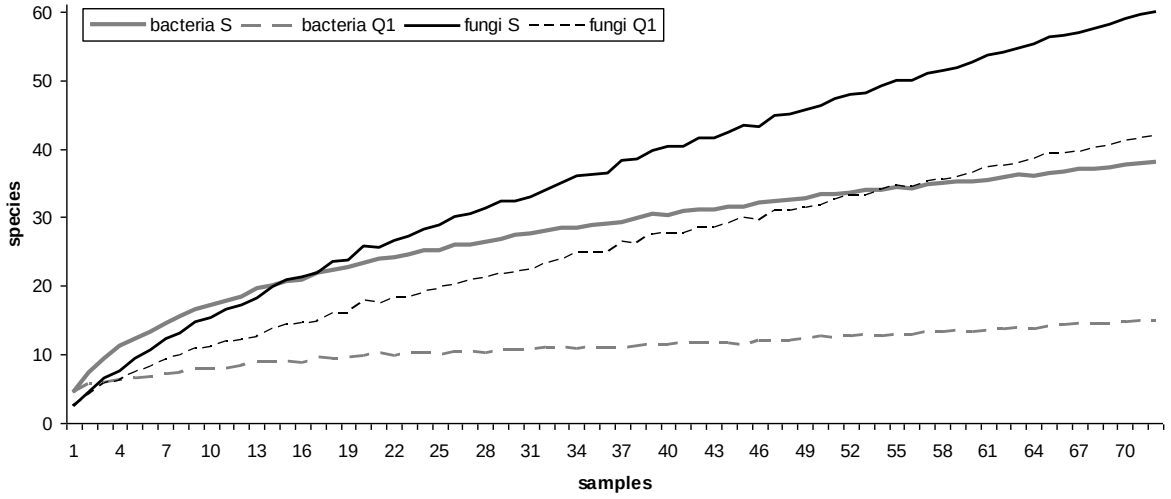


Figure 3: Species accumulation curves are shown for bacterial and fungal OTUs detected in 72 sputum samples. Averaged curves show the number of detected species versus number of samples. Curves for bacteria are shown in grey and curves for fungi are shown in black. Full lines show observed species counts (S; Mao Tao) and dashed lines show unique species counts (Q1; species which occur in only one sample).

Overall, contrary trends were observed for fungal and bacterial species detected in the CF cohort. Whereas, the number of bacterial counts seemed to stabilize beyond 72 samples, new

fungal species have to be continuously considered with increasing sample numbers. Species accumulation curves for 56 sputum samples (only the first collected sample from each patient) were likewise calculated and confirmed these results (**Figure S3**). Therefore, a high richness of fungal species in contrast to a relatively small number of bacteria was determined in this CF cohort study.

From a subgroup of 13 CF patients, sputum samples were collected twice or three times over a period of 2 years. Comparing the individual species richness, this subgroup also showed a different picture for fungi and bacteria (**Figure 4**). For fungi, the star-like appearances of the curves indicate major differences in numbers of observed OTUs. Individual numbers are rather diverse throughout the 13 patients as well as between the different time points of the same patient. In contrast, the comparatively circular-like appearances of the curves for bacteria indicate a relatively stable number of OTUs throughout the patients as well as between the different time points of the same patient. One exception was patient #7: samples showed at the first time point of collection only 2 bacterial OTUs whereas in the second sputum sample 10 OTUs were detected. Further correlations between the individual numbers of fungal and bacterial OTUs for one patient could not be observed.

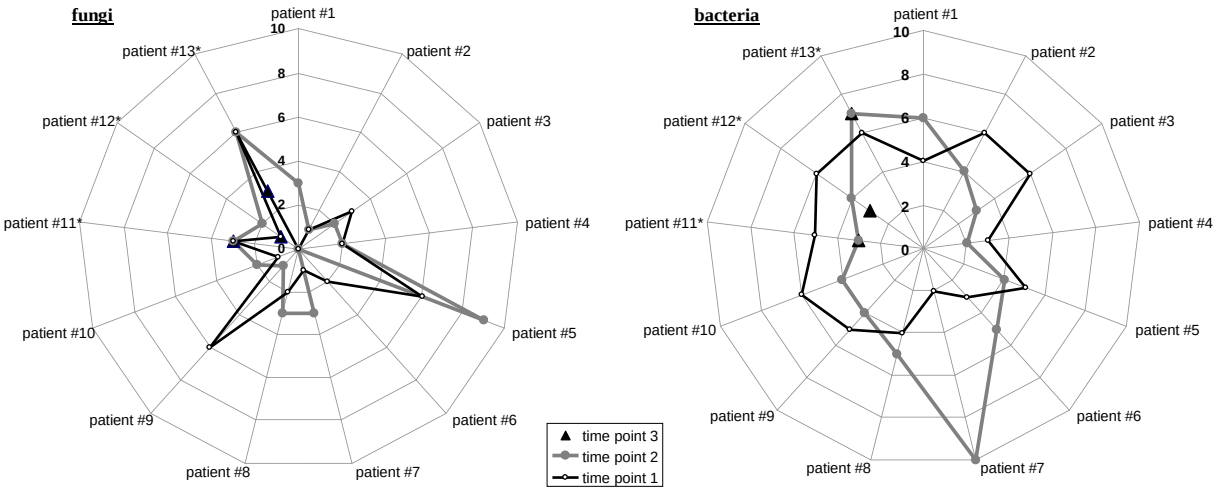


Figure 4: Differences in individual bacterial and fungal OTU counts for a subgroup of 13 CF patients with repeated sputum collection. Radar charts are showing patient #1 to #13 on angular axes. Numbers of detected species are given in radial axes. Blank and grey symbols indicate values for first and second sputum samples, respectively. Symbols are accordingly combined with black and grey lines to locate the correct values for each patient. Three sputum samples were analyzed for patients #11*-13* and values are represented by black triangles only. Fungal OTU counts are shown in the left and bacterial counts in the right plot.

3.4.2 Species diversity and occurrence frequency of bacteria in CF cohort

Bacterial species from nine major taxonomic groups at the class level were observed in the sputum samples of the CF cohort (**Figure 5**). The highest number of positive samples (occurrence) was observed for the 14 OTUs of the class *Bacilli*, including *Streptococcus salivarius* and *Streptococcus parasanguinis* which were both detected in 56.9% of the samples (**Table S1)**. Considering the individual occurrences of all OTUs, all other taxonomic classes, with the exception of *Bacteroidia*, were dominated by a single OTU. In 70.8% of the samples *Gammaproteobacteria* were observed and *P. aeruginosa* was detected in 69.4%. *Actinobacteria* with 6 detected species were dominated by *Rothia mucilaginosa* which was found in 43% of the samples. Three species of *Betaproteobacteria* were considered, *Achromobacter xylosoxidans* was the most frequent with 8.3% positive samples. Four taxonomic classes with a single OTU each were observed, *Fusobacterium nucleatum* (*Fusobacteria*) was detected in 6.9% of the samples and *Veillonella* -1 (*Negativicutes*) in 5.5%. Unique classes were found in just one sample and were represented by *Capnocytophaga gingivalis* (*Flavobacteria*) and an oral TM7 candidate taxon. Further details on individual occurrences are given in **Table S1**.

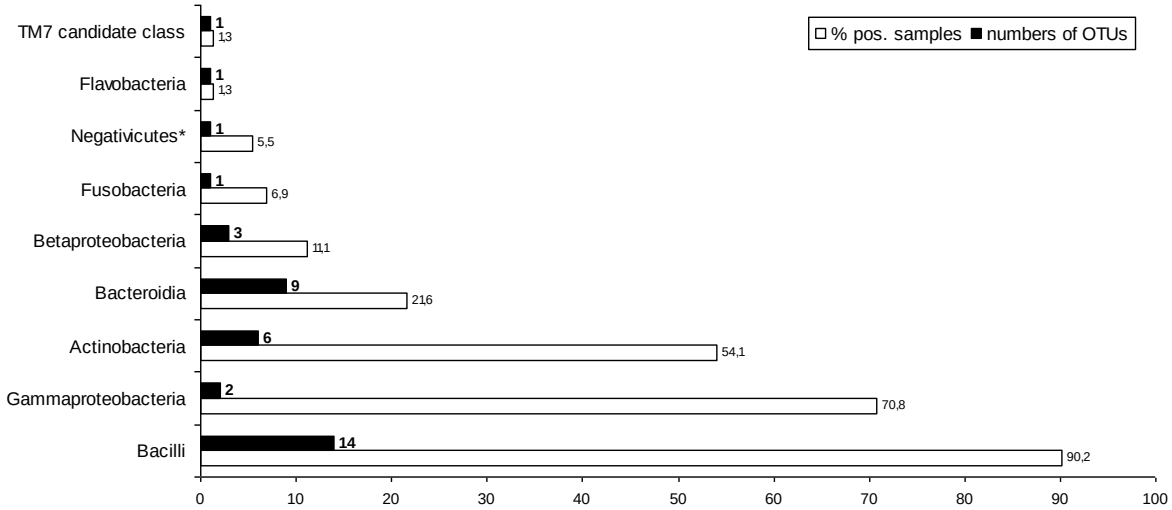


Figure 5: Frequency of taxonomic classes of bacteria detected in 72 CF sputum samples. Black bars give number of OTUs observed for each class. Grey bars indicate percentage of positive samples of respective OTUs. Precise values for each bar are given in the figure.

*formerly member of the class *Clostridia*

In the class *Bacteroidia*, 9 OTUs of the genus *Prevotella* were observed. *P. histicola* and *P. melaninogenica* were the most frequently detected with 6.9% positive samples each. The phylogenetic diversity of all observed bacterial species was depicted in **Figure 6.** Likewise, the genus *Streptococcus* was represented by 9 OTUs in the CF cohort. Most of them are

considered to be members of the *Streptococci mitis* group with the exception of *S. salivarius* and two species of the *S. millerii* group, *S. anginosus* and *S. intermedius* which were detected in 8.3% and 6.9% of the samples, respectively. For the OTU affiliated to *S. mitis* itself an occurrence rate of 43% was observed.

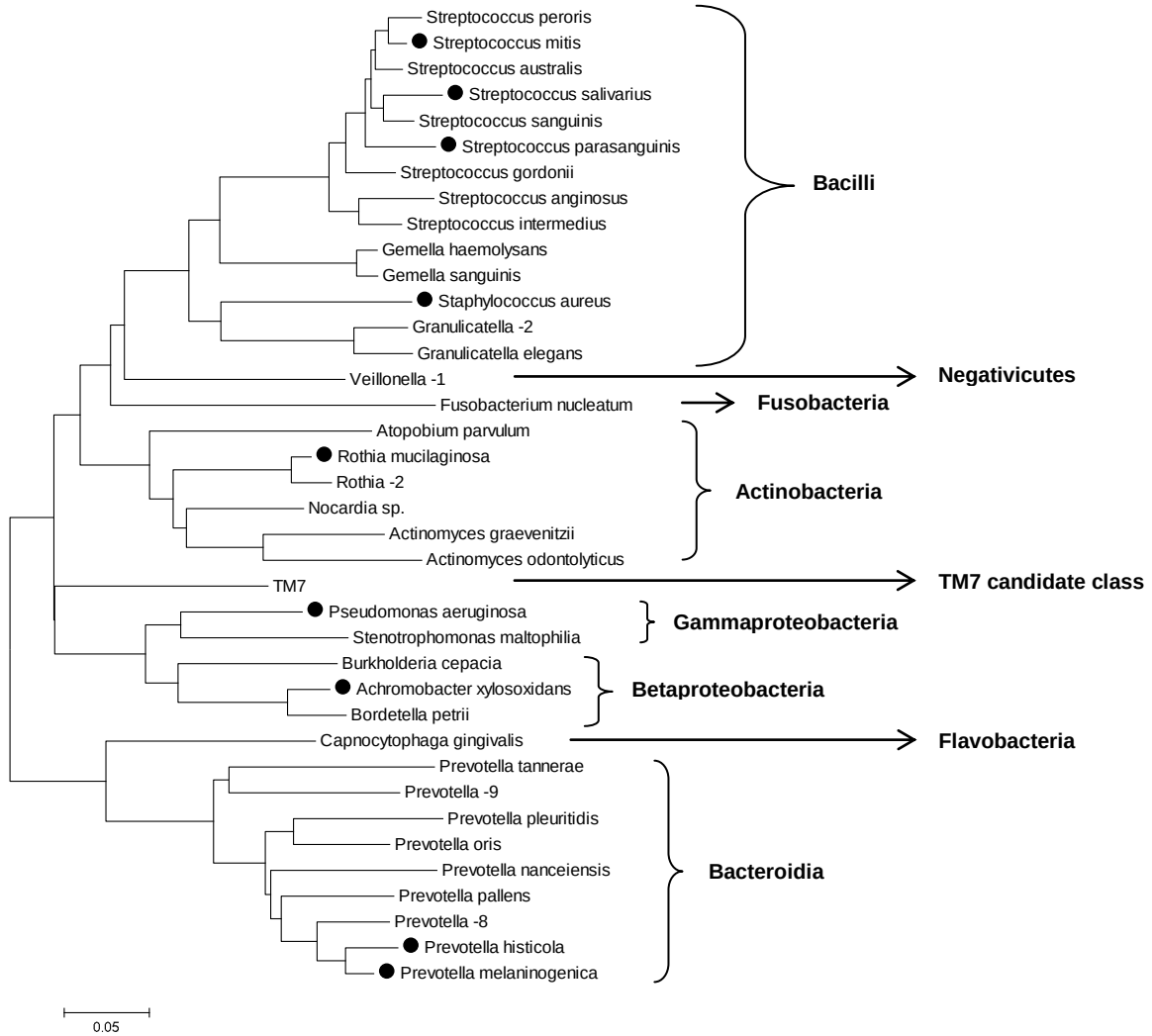


Figure 6: Phylogenetic tree for all bacterial OTUs observed in the 72 sputum samples based on 16S rRNA gene sequence analysis. Bacteria from 8 taxonomic classes were observed, representative sequences from closest described bacterial species for each OTU were used to construct the tree. Dominant species of each class are additionally labelled with black circles. Scale bar represents base substitution per site.

From the class *Bacilli*, three *Staphylococcaceae* species were found, including *S. aureus*, with 25% positive samples the most frequent non-Streptococci, and two *Gemella* species. *G. sanguinis* was found in 20.8% and *G. haemolysans* in 11.1% samples. From *Actinobacteria*, two OTUs each in the genus *Rothia* and the genus *Actinomyces* were detected. For *Actinomyces*, *A. graevenitzii* was the most frequent with 13.8% positive samples. Overall, five species from the phylum *Proteobacteria* were observed, including *S. maltophilia* (*Gammaproteobacteria*) detected in 4.1% of the samples and *B. cepacia* as well as *Bordetella*

petrii (both *Betaproteobacteria*) were detected in 1.3%. All OTUs not particularly mentioned in the text were found in less than 5 samples (detailed individual occurrences are mentioned in **Table S1**). Absolute number of incidences for each bacterial class and an overview of the occurrences of the bacterial species are given in **Table 1.a**.

3.4.3 Species diversity and occurrence frequency of fungi in CF cohort

Ten taxonomic classes could be identified from the two fungal divisions *Ascomycota* and *Basidiomycota*. One OTU could not be taxonomically assigned, but partial 18S and 5.8S sequences of the amplicon exhibit typical fungal structures and was therefore considered. Additionally, one OTU in each division did not exhibit sufficient similarity to be further assigned to a specific class. In total, members of 13 fungal classes were observed (**Figure 7**). While for bacteria a single dominant species was found in most taxonomic classes, no such observation was made for fungi.

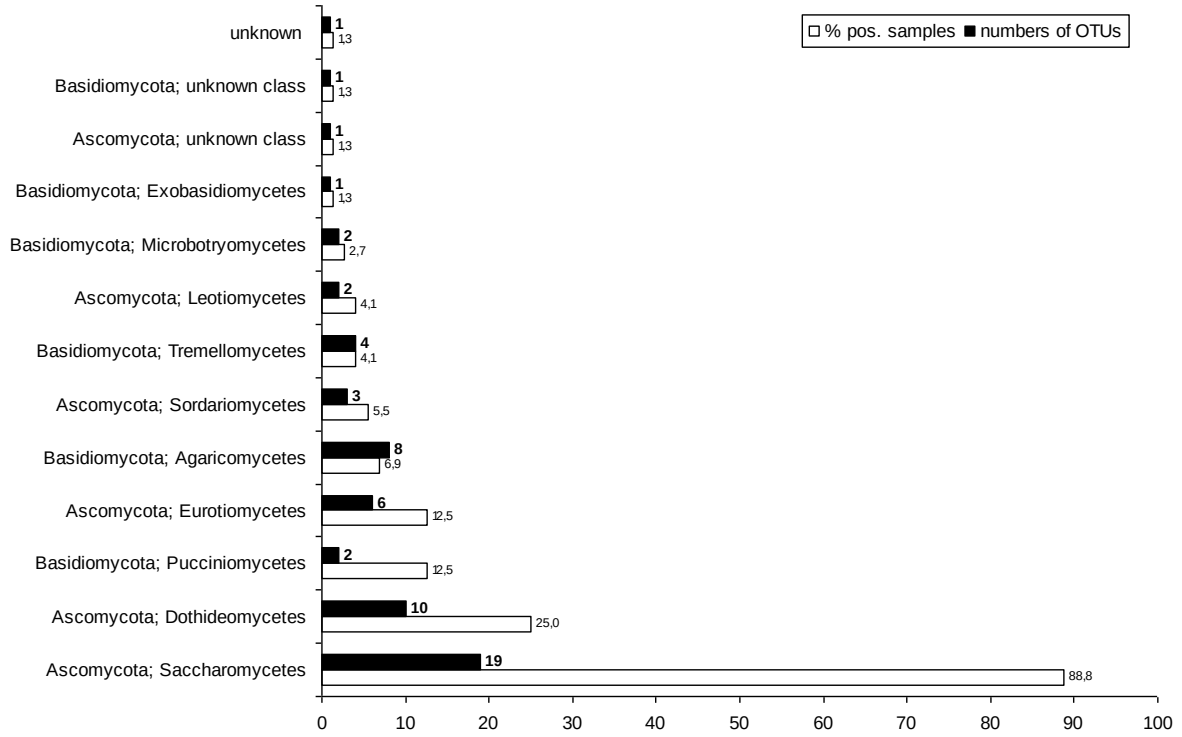


Figure 7: Frequency of taxonomic classes of fungi detected in 72 CF sputum samples. Black bars give number of OTUs considered for each class. Grey bars indicate incidence of associated OTUs. Values for each bar are given in the graphic.

Saccharomycetes were the most frequently found fungi and this class comprised the five most frequently found species. 88.8% of samples were positive for the 19 OTUs observed. The polymorphic yeasts *C. albicans* and *C. dubliniensis* were the dominant fungal species in the CF cohort with 44.4% and 23.6% positive samples respectively, followed by *S. cerevisiae*

with 19.4% and *Candida parapsilosis* with 13.8% positive samples. *C. glabrata* was detected in 12.5%. From *Dothideomycetes*, 10 species in 25% of the samples were observed. Two OTUs were dominant in this class: the *Cladosporium cladoporoides* complex with 11.1% and the *Cladosporium herbarum* complex with 9.7%. The basidiomycete class *Pucciniomycetes* showed only two species, the yeasts *Sporobolomyces roseus* and *Sporobolomyces ruberrimus* which were found in 11.1% and 5.5% of sputum samples, respectively. These two species were repeatedly detected together in the same samples (n=3). Within the other taxonomic classes, individual incidences of the associated OTUs were more equally distributed. An overview about the occurrences of major fungal species is given in **Table 1.b** and detailed individual occurrences of fungal OTUs are mentioned in **Table S2**

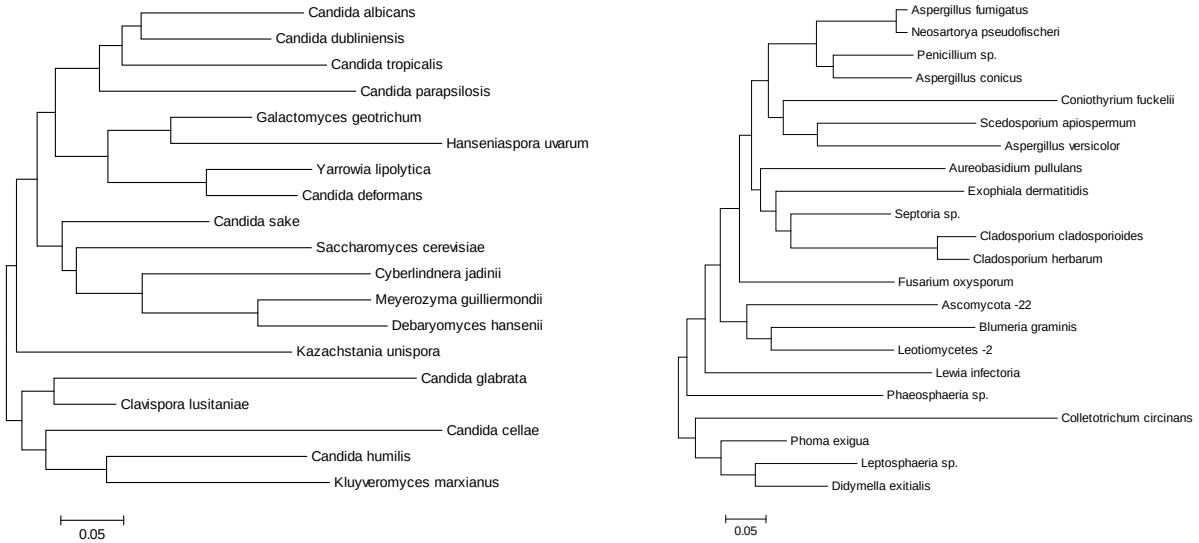


Figure 8: Phylogentic trees for fungal OTUs observed from the division *Ascomycota* based on ITS region sequence comparisons. On the left: *Saccharomycetes* species found in the CF cohort. *C. albicans* and *C. dubliniensis* were the most frequently found fungi in the CF cohort. On the right: non-Saccharomycetes OTUs including filamentous fungi considered as major CF pathogens like *Aspergillus spp.*, *Exophiala dermatitidis* and *Scedosporium apiospermum* (teleomorph: *Pseudoallescheria boydii*). Scale bar represents base substitution per site.

From the total of 60 fungal OTUs observed, 41 could be assigned to the division *Ascomycota* and 18 to the division *Basidiomycota*. Due to their high diversity, ascomycetes OTUs were further split into *Saccharomycetes* species and non-*Saccharomycetes* (**Figure 8**). Likewise, all OTUs considered *Basidiomycota* were depicted in a third phylogenetic tree (**Figure 9**). Several filamentous mould species from the classes *Dothiomycetes* and *Eurotiomycetes*, including four *Aspergillus*-like species (*A. fumigatus*, *A. conicus*, *A. versicolor* and *Neosartorya pseudofischeri*), were observed in the CF cohort and represent most of the OTUs in the associated phylogenetic tree. From the basidiocarp-forming *Agaricomycetes*, eight

mushroom-producing fungi were observed, like *Plicaturopsis crispa*, *Megacollybia platyphylla*, *Piptoporus betulinus*, *Polyporus gayanus*, *Strobilurus sp.*, *Baeospora sp.* and *Hyphodontia sp.*.

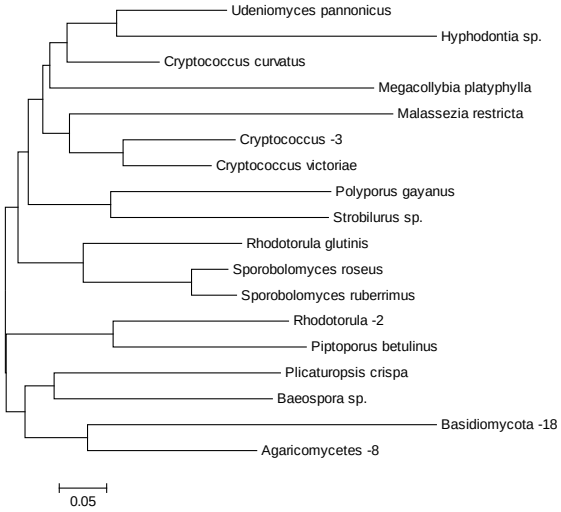


Figure 9: Phylogentic tree for fungal OTUs observed from the division *Basidiomycota* based on ITS region sequence comparisons. Species from several basidiomycetes classes were detected. Scale bar represents base substitution per site.

Overall, a high species diversity of fungi was observed in sputum samples from the CF cohort. Both major fungal divisions were represented by several classes and species. Yeast-like fungi were found in all three groups and showed the highest diversity. Poly- and dimorphic fungal species were found as well as species limited to a single type of growth.

3.4.4 Dynamics of fungal OTUs and clinical abnormalities in individual CF patients

In **Figure 4** differences of individual fungal OTU counts for a subgroup of 13 CF patients are presented. In addition to the total number of fungi, the individual incidences of single OTUs over time were analyzed in this subgroup and revealed rare OTUs that were repeatedly associated with the same patients. These rare OTUs were further characterized by overall low occurrence rates. The emerging CF pathogen *Exophiala dermatitidis* from the class *Eurotiomycetes* was found in two samples collected within a three-month period from patient #3. With a total of two and three fungal OTU counts, a relatively constant number of species was observed for this patient. The major CF pathogen *Scedosporium apiospermum* (*Sordariomycetes*) was detected in two sputa samples collected within a three-month period from patient #5. In total, six fungal species were observed in the first and nine OTUs in the second sputum sample. In the second, several filamentous fungi were found like *Penicillium* sp. (*Eurotiomycetes*), *Fusarium oxysporum* (*Sordariomycetes*) and both *Cladosporium* OTUs mentioned above, whereas *Aspergillus fumigatus* (*Eurotiomycetes*) was only detected in the first sample of that patient. Two *Rhodotorula* OTUs, from the class *Microbotryomycetes*, were observed in the CF cohort. Both of these basidiomycetes yeast-like species were found

in patient #13* showing an advanced number of 6 fungal species counts in the sputum. A second sample from this patient collected 15 months later still exhibited 6 fungal OTUs, though comprising mostly yeast and not *Rhodoturola* species OTUs. Five months later, a third sample from patient #13* showed a reduced number of species but the major CF pathogen *Scedosporium apiospermum* was found.

Clear correlations between single fungal OTUs and clinical features of CF patients could not be observed. However, from five CF patients exhibiting an unusually high number of individual fungal OTU counts (>4) three showed conspicuous clinical abnormalities. As mentioned above, in two sputum samples from patient #5 six and nine fungal OTUs were observed. ABPA was diagnosed for this patient. From the same subgroup, patient #13* exhibited 6 fungal OTUs in the first two sputum samples and lung function tests revealed low FEV_1 values (predicted forced expiratory volume in one second), not exceeding 30% at any time of sputum collection. In total three *Cryptococcus* OTUs (*Tremellomycetes*) were observed in the CF cohort, two of these yeasts were detected in a sample of an additional patient who exhibited an unusually high number of 7 fungal OTUs in the sputum. This patient was suffering from an acute infection with advanced levels of inflammatory markers in the blood (C-reactive protein: 129 mg/l).

Four fungi from the class *Agaricomycetes*, which comprises typical mushrooms, were found in one sample, together with the maximum number of eleven fungal OTU counts. Although no conspicuous clinical abnormalities were mentioned, this patient was characterized by an untypical CF caused by a mutation in the sodium channel. Besides the four cases described above, one additional patient exhibited an unusually high number of fungal OTUs in the sputum. No conspicuous clinical abnormalities were observed. Six fungi in total were detected of which four belong to the class *Saccharomycetes*, all OTUs were repeatedly observed throughout the CF cohort with an occurrence rate of at least 4.1%.

1.a	phylum	class (total no. of incidences)	major species	% pos. samples
Bacteria	Firmicutes	Bacilli (185)	Streptococcus salivarius	56.9
			Streptococcus parasanguis	56.9
			Streptococcus mitis	43.0
			Staphylococcus aureus	25.0
			Gemella sanguinis	20.8
			Granulicatella -2	19.4
			Gemella haemolysans	11.1
			Streptococcus anginosus	8.3
			Streptococcus intermedius	6.9
		Negativicutes (4)	Veillonella -1	5.5
	Proteobacteria	Gammaproteobacteria (53)	Pseudomonas aeruginosa	69.4
			Stenotrophomonas maltophilia	4.1
		Betaproteobacteria (8)	Achromobacter xylosoxidans	8.3
			Bordetella petrii	1.3
			Burkholderia cepacia	1.3
	Actinobacteria	Actinobacteria (50)	Rothia mucilaginosa	43.0
			Actinomyces graevenitzii	13.8
	Bacteroidetes	Bacteroidia (21)	Prevotella histicola	6.9
			Prevotella melaninogenica	6.9
		Flavobacteria (1)	Capnocytophaga gingivalis	1.3
	Fusobacteria	Fusobacteria (5)	Fusobacterium nucleatum	6.9
	TM7 candidate phylum	TM7 candidate class (1)	TM7	1.3

1.b	division	class (total no. of incidences)	major species	% pos. samples
Fungi	Ascomycota	Saccharomycetes (101)	Candida albicans	44.4
			Candida dubliniensis	23.6
			Saccharomyces cerevisiae	19.4
			Candida parapsilosis	13.8
			Candida glabrata	12.5
			Candida tropicalis	4.1
			Cyberlindnera jadinii	4.1
			Candida sake	2.7
		Dothideomycetes (24)	Cladosporium cladosporioides	11.1
			Cladosporium herbarum	9.7
			Lewia infectoria	2.7
		Eurotiomycetes (9)	Exophiala dermatitidis	4.1
			Aspergillus fumigatus	2.7
		Sordariomycetes (5)	Scedosporium apiospermum	4.1
		Leotiomycetes (3)	Blumeria graminis	2.7
	Basidiomycota	Pucciniomycetes (12)	Sporobolomyces roseus	11.1
			Sporobolomyces ruberrimus	5.5
		Microbotryomycetes (3)	Rhodotorula glutinis	2.7

Table 1.a, b: Occurrence of major OTUs detected in 72 CF sputum samples. Table **1.a** comprises bacterial OTUs described in the text. Table **1.b** shows fungal OTUs detected more than once. Bacterial and fungal OTUs are grouped according to taxonomic classification and ranked according to number of positive samples (in %). Additionally, the absolute number of detected incidences for each class is given in brackets (considering all detected OTUs including unique species). The following basidiomycetes classes are represented by unique OTUs only and are therefore not considered in the table: *Agaricomycetes*, *Tremellomycetes*, *Exobasidiomycetes*.

3.5 Discussion

3.5.1 Microbial species identification

The analysis of microbial species from the sputum samples was performed using ribosomal RNA gene fragments and sequences: For bacteria the V1-V3 region of the 16S rRNA gene was used and for fungi the internal transcribed spacer (ITS) between the 18S rRNA and 5.8S rRNA gene. Elucidation of bacterial communities in CF sputa by different techniques of 16S rRNA gene analysis is the most commonly used culture-independent strategy and shown to reveal more species associated with the respiratory tracts than cultivation of bacteria (Rogers et al. 2003) (Guss et al. 2011). Next generation sequencing methods are capable of detecting species down to lower relative abundances than classical molecular fingerprints. However, the clinical relevance of very low abundant bacteria within a community is disputable. In the best case, SSCP electrophoresis is considered to have a detection limit down to a relative abundance of 0.1% (Eichler et al. 2006). In the current study, a detection limit of about 1% was assumed. Furthermore, the variable regions targeted were shown to be appropriate for the differentiation of bacterial pathogens in diagnostics (Chakravorty et al. 2007).

Besides the bacterial microflora, fungal communities were also detected with the same strategy. The fungal target region is further known to exhibit high length heterogeneity across fungal species. The species-specific length polymorphism was used in previous studies to characterize fungal communities and to identify medically important fungi (Chen et al. 2001) (Ranjard et al. 2001). SSCP electrophoresis was successfully applied for the diagnosis of fungal infections in other body sites (Kumar & Shukla 2005). Furthermore, the target region chosen in the current study was shown to be appropriate for community representation and species resolution (Schoch et al. 2012). Overall, the identification and detection methods chosen for the current study seemed to be appropriate to obtain a comparable level of taxonomic resolution, i.e. species of bacteria and fungi. High taxonomic informativeness and sequence similarity allowed confident identification of the species for most OTUs. It is important to note here, that the methods are DNA-based and therefore no distinction is made between respiratory tract microbial colonisers, DNA originating from dead bacteria, inhaled fungal spores or tissue fragments.

3.5.2 Bacterial species present in CF sputum

The species diversity of bacteria in the CF cohort was described here to be comparatively low. Individual communities were partly highly diverse and rather different from each other but the more sputum samples that were analyzed, the more known (previously observed) bacterial species were detected. Similarly, individual OTU counts showed a comparatively constant number of species throughout different patients and between different sputum samples of the same patients. Higher bacterial richness found in other CF cohort studies resulted from detection sensitivities down to 0.012% of the total volume and less than 1% volume for the majority of observed OTUs (Rogers et al. 2004). Taking only higher abundant species into account, the current study presents the major bacterial taxa of a broad CF cohort. The concept of defining core taxa in bacterial communities of CF airways was approached previously by Gast et al. (Gast et al. 2011). This study divided bacterial taxa from 14 clinically stable CF patients into a core and a satellite group, the core comprised only 15 taxa and the satellite group included major CF pathogens. The current study suggests a core group consisting of a broader number of bacterial taxa: 23 bacteria from 7 taxonomic classes were detected more than once. These bacteria comprised the major CF pathogens *P. aeruginosa*, *S. aureus*, *S. maltophilia* and candidate CF pathogen *A. xylosoxidans* (LiPuma 2010) (Hauser et al. 2011). However, the important pathogen *B. cepacia* was only detected in one sample. Together with several other rarely reported candidate CF pathogen in the literature, like *Nocardia* sp., the number of CF relevant taxa is beyond the 23 described in the current study.

Concerning the origin of bacteria present in CF sputum, the communities in the human oral cavity has to be considered. Most of the observed bacterial species in the current study are part of the common oral micoflora (Zaura et al. 2009). However, some of the most abundant oral genera were not detected at all in the current study, suggesting a different composition of the bacterial communities detected in the sputum. The oral cavities are assumed to potentially act as a reservoir and 'stepping stone' for bacterial immigration into the CF lung (Rogers et al. 2006). As a consequence, oral bacteria might substantially influence the microbial community compositions in the lower airway and their presence in sputum should not only be seen as contamination from oral cavities. The healthy oral microflora recently presented by the human microbiome project, revealed the genera *Streptococcus*, *Haemophilus*, *Actinomyces* and *Prevotella* to be the most dominant taxa. Abundances of species in the genus *Streptococcus* were further elucidated: *S. parasanguinis* and *S. salivarius* were the most dominant species, whereas *S. mitis* was only ranked as the sixth most abundant one and was comparable to *S.*

peroris (The Human Microbiome Project Consortium 2012). Both dominant species were also observed to be the most prevalent ones in the CF cohort. But *S. mitis* was detected in 43% of the samples and therefore the third most common *Streptococcus* species, whereas *S. peroris* was detected in less than 5 sputa. Overall, the similarities and main differences in communities present in oral cavities and CF sputum samples suggests that the upper and deeper respiratory tract microflora are closely related but selective conditions promote growth of different species resulting in bacterial communities with altered frequencies and abundances in CF sputum.

A high similarity was observed between the dominant OTUs in the current study and highly abundant OTUs observed in two recent studies, one investigating the bacterial community dynamics over one decade in six CF patients and the other analyzing the phylogenetic diversity of the human respiratory microbial ecosystem in 4 children with CF (Guss et al. 2010) (Zhao et al. 2012). These similarities support the recent observation of a correlation between the occurrence frequency of OTUs and their relative abundances (Gast et al. 2011) (Zhao et al. 2012). The frequent detection of the same dominant bacterial species throughout different studies suggests growth advantages for these species in the CF respiratory tract and therefore the communities in CF airways are presumably highly influenced by the dominant species, like *P. aeruginosa*, *S. aureus*, *Rothia mucilaginosa* and *Streptococcus spp.*

The clinical relevance of the bacterial species present in sputum is a key issue in microbiological studies of CF patients. End-stage lungs are shown to harbour relatively low numbers of species and 16S rRNA gene analysis of explanted lung tissue samples revealed the same four phyla of the five bacterial classes with more than one OTU found in the current study: *Proteobacteria*, *Bacteroidia*, *Firmicutes*, *Actinobacteria* (Rudkjøbing et al. 2011). Rudkjøbing et al. (Rudkjøbing et al. 2011), further reported the same bacterial species to be found in sputum taken prior to the lung explantation. A notably high species diversity for the families *Prevotellaceae* and *Streptococcaceae* was recently described similar to the current study and particularly the pathogenicity of the *S. millerii* group is under discussion (Sibley et al. 2008) (Sibley et al. 2011). *S. millerii* species, frequently found in the current study, is suggested to be associated with a decline in health. A consistent correlation of these *Streptococci* and co-colonisation with *P. aeruginosa* was observed and a resultant enhanced virulence of the bacteria suggested (Parkins et al. 2008). Several particularly anaerobic bacterial species were found, that may potentially be CF pathogens, like *Prevotella spp.* and

Actinomyces spp. (LiPuma 2010). More studies are needed to compare their occurrence and relative abundances with a decline in the health condition of CF patients.

3.5.3 Fungal species present in CF sputum

The richness of fungal species in the CF cohort was shown to be high. The large number of unique fungal OTUs observed only once implies that the actual species richness of fungi is presumably even larger than detected. Individual fungal OTU counts showed a rather diverse picture throughout patients and between different sputum samples of the same patient. Therefore, a high fluctuation of fungi present in CF airways has to be assumed.

The question about the origin of fungal DNA arises. Aerial dispersal is considered to be a major distribution factor of many fungal species and the high richness of fungal species was recently analyzed in particulate matter of air from the outskirts of a major city in Germany (Fröhlich-Nowoisky et al. 2009). Similarly, a high number of unique species observed only once was also found in air and more importantly all fungal classes identified in the current study were also observed by Fröhlich-Nowoisky et al. (Fröhlich-Nowoisky et al. 2009) with a relative proportion of at least 1% in the air. Furthermore, some classes of a major proportion of the detectable fungi in air like *Dothideomycetes*, *Eurotiomycetes*, showed similar occurrence frequency in the clinical samples, suggesting that the fungal species diversity in air is mirrored in sputum. Interestingly, the dominant fungi in air and in sputum showed a contrary picture. The most frequently detected and most diverse fungal class in CF sputum was *Saccharomycetes*, which represents only a small minority in the air. In contrast, *Agaricomycetes* were shown to be the dominant class in air but only 6.9% of sputum samples were positive (Fröhlich-Nowoisky et al. 2009). These comparisons between fungal occurrences in the major exposure medium (air) and sputum samples suggest that fungi present in CF airways are not only a reflection of the species in the ambient air. Once they are inhaled, selective conditions presumably promote rapid removal or persistence of fungal species. Mushroom-forming fungi and several plant pathogens are more likely to be removed quickly, as seen for *Agaricomycetes* or *Leotiomycetes* species, like *Blumeria graminis* (powdery mildew). In contrast, yeast species are more likely to find conditions to temporally colonize the human respiratory tract in CF patients. Accordingly, the most frequently detected fungi from both fungal divisions are yeast-like species: *Pucciniomycetes*, represented by two *Sporobolomyces* species, and *Saccharomycetes*, dominated by *Candida* species.

Several fungi from the observed classes are yeast-like species. Their ability to persist in the respiratory tract can be elucidated by comparisons of the dynamics of fungal OTUs in sputum. Unlike *Candida* species, *Sporobolomyces* species are among the most abundant fungi of indoor air (Fairs et al. 2010). Half of all incidences were detected in only three sputum samples from two patients. Together with *S. roseus*, also *Cryptococcus* and *Rhodotorula* species are considered to be the most important viable yeasts in outdoor air, though generally only about 2% of all viable spore counts are yeast species (Rantio-Lehtimäki 1985). Again, species from these genera were rarely found but repeatedly detected in the same samples, suggesting certain conditions in the airways of those patients which are promoting persistence of these yeast species, whereas they are removed quickly from the airways of other individuals.

Though, the genus *Candida* is almost not present in air, these fungi are known to represent the most frequent species in the oral fungal microbiome (Ghannoum et al. 2010). Several studies confirmed *C. albicans* to be the most isolated fungi from CF patients (Bakare et al. 2003) (Valenza et al. 2008) (Muthig et al. 2010). The second most commonly isolated fungus in the current study was *C. dubliniensis* with 23.6% positive samples. This yeast is of particular interest: not much is known about its prevalence in other body sites or the environment but colonization rates over 10% where also observed in other CF studies (Peltroche-Llacsahuanga et al. 2002) (Muthig et al. 2010). In terms of morphology and genetics, this fungus is closely related to *C. albicans* though several important virulence genes are missing (Moran et al. 2012). *C. dubliniensis* is rarely found in non-CF individuals, which might be explained by its ability to display cell surface hydrophobicity-associated adherence at 37°C that enables the fungus to take advantage of the dehydrated respiratory secretion of CF patients (Hazen et al. 2001) (Peltroche-Llacsahuanga et al. 2002). The other important *Candida* species, namely *C. parapsilosis*, *C. glabrata* and *C. tropicalis* are observed with lower occurrence rates but different CF studies showed a surprisingly similar ranking of occurrence (Doern & Brogden-Torres 1992) (Müller & Seidler 2010). Therefore, *Candida* species are presumably a common colonizer of airways in CF patients. This high conformity suggests a similar picture observed for the dominant bacterial taxa. Conditions in the CF airways might differentially promote growth of these species according to their rank of occurrence, resulting in this prevalently observed pattern for the genus *Candida*. In this context, corresponding individual relative abundances of species in the fungal communities would be expected and therefore the

communities in CF airways are presumably highly influenced by these dominant *Candida* species.

3.5.4 Clinical relevance of fungal species

Despite the high isolation rates in the sputum, the virulence of *Candida* spp. in the context of CF respiratory tract infections is under debate (LiPuma 2010). Mainly considered to be clinically relevant causative fungal agents of respiratory tract infections in CF patients are filamentous fungi from classes *Sordariomycetes* and *Eurotiomycetes*, namely *A. fumigatus*, *S. apiospermum* (teleomorph: *Pseudallescheria boydii*) and *E. dermatitidis* (in Germany) (Müller & Seidler 2010) (LiPuma 2010). The occurrence rates of these fungi vary strongly in different studies and the current cohort showed comparatively low rates, especially with 2,7% for *A. fumigatus* (Pihet et al. 2009). Nevertheless, positive samples exhibited strong distinctions of clinical relevance: *A. fumigatus* was detected in the only patient with a diagnosted ABPA, likewise *S. apiospermum* was repeatedly detected in the same patient. The latter one was accordingly recently described to be associated with *A. fumigatus*, usually following its colonization of the CF airways (Pihet et al. 2009). The emerging pathogen *E. dermatitidis* was repeatedly found in sputum from the same patient within 3 months, suggesting persistence of this fungus in the patient. The virulence of a few other filamentous fungi is further under debate, including the highly prevalent *Cladosporium* spp. which are similarly found at high frequency in air and in healthy oral cavities (Fröhlich-Nowoisky et al. 2009) (Ghannoum et al. 2010). However, confirmed reports about colonization of human airways by these species are missing.

In consideration of the high abundances of fungal spores in indoor and outdoor air and the high phylogenetic diversity of fungal species in CF sputum, the number of reported and confirmed cases of fungal respiratory infections is relatively low. Just a very few species from these highly diverse airborne contaminants are considered to be the causative agents of infections in CF, and even these fungi are often considered to be commonly found in the environment, suggesting a high exposure of CF patients to these fungal spores. Therefore, the current study suggests that virulence of fungal species in the context of human respiratory infections should be considered to be generally low and probably depending on further circumstances which remain to be elucidated.

An advanced number of fungal species in the respiratory tract, however, may act as an indicator for changing conditions associated with bad health. Patients with advanced fungal species numbers often showed conspicuous clinical abnormalities. For example, the sputum from the patient with the ABPA exhibited advanced numbers of filamentous fungal species in both sputum samples. The highest C-reactive protein level in the CF cohort was measured in one patient with seven fungal species in the sputum. In a third patient, high species counts were further measured in two sputa over a long time period (15 month) which showed constantly low FEV_1 values and subsequently the same patient exhibited the fungal CF pathogen *S. apiospermum* another 5 months later.

3.6 Conclusion & Outlook

Comparing the fungal and bacterial richness associated with CF sputum, bacterial communities consist of a relatively small number of species. Respiratory tract communities are closely related to the oral microbiome but significant differences were observed. A core group of major bacterial taxa associated with CF is presented. A high conformity of these species with other CF studies was observed, suggesting advanced colonization abilities for these bacteria. Although the clinical relevance for most of the species is unclear, a major influence on the pulmonary microbial community for the dominant bacteria is assumed. A potential influence of these species on the pathogenicity of the entire community remains to be elucidated as well as their role as single pathogenic species.

A high richness for fungal species was determined for the cohort and high fluctuations of different species were observed. The frequencies of occurrence of fungi in the air as a major exposure medium showed similarities, but major differences were observed for the dominant classes. Indications for individual persistence of yeast-like fungi in respiratory tracts were shown and high colonization abilities for given *Candida* species assumed. Although major fungal pathogens were repeatedly observed in sputum samples, the overall virulence of fungal species was considered to be low. However, advanced numbers of fungi in the respiratory tract may act as an early indicator for a progression towards infection-related illness. High exposure rates in contrast to low occurrence rates imply that colonization and persistence of filamentous fungi is presumably related to certain ecological factors and depending on circumstances which promote these rare events. Therefore, fungal infections may substantially be reduced if these ecological interactions are better understood.

Microbial growth and persistence promoting conditions have to be elucidated and further studies on the ecology of pulmonary communities are necessary, combining the bacterial and fungal microbiota. Hereby, interactions of the dominant bacteria and yeasts presented in the current study may have a major impact on the respiratory microbial communities. A deeper understanding of ecological factors that promote the colonization by fungi and increase the pathogenicity of bacterial communities may substantially improve the diagnostics and prediction in the progression of respiratory diseases.

3.7 References

Bakare N, Rickerts V, Bargon J & Just-Nübling G (2003) Prevalence of Aspergillus fumigatus and other fungal species in the sputum of adult patients with cystic fibrosis. *Mycoses* 46: 19–23.

Bassam BJ, Caetano-Anollés G & Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80–83.

Boucher RC (2004) New concepts of the pathogenesis of cystic fibrosis lung disease. *Eur. Respir. J.* 23: 146–158.

Chakravorty S, Helb D, Burday M, Connell N & Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69: 330–339.

Chen Y-C, Eisner JD, Kattar MM, Rassoulian-Barrett SL, Lafe K, Bui U, Limaye AP & Cookson BT (2001) Polymorphic Internal Transcribed Spacer Region 1 DNA Sequences Identify Medically Important Yeasts. *J Clin Microbiol* 39: 4042–4051.

Cole JR et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37: D141–145.

Cugini C, Calfee MW, Farrow JM, Morales DK, Pesci EC & Hogan DA (2007) Farnesol, a common sesquiterpene, inhibits PQS production in Pseudomonas aeruginosa. *Mol. Microbiol* 65: 896–906.

Cystic Fibrosis Foundation Patient Registry (2012) *2011 Annual Data Report*. Cystic Fibrosis Foundation, Bethesda, Maryland http://www.cff.org/UploadedFiles/research/ClinicalResearch/2011-Patient-Registry.pdf (Accessed January 8, 2013).

Delhaes L et al. (2012) The Airway Microbiota in Cystic Fibrosis: A Complex Fungal and Bacterial Community—Implications for Therapeutic Management. *PLoS ONE* 7: e36313.

Doern GV & Brogden-Torres B (1992) Optimum use of selective plated media in primary processing of respiratory tract specimens from patients with cystic fibrosis. *J Clin Microbiol* 30: 2740–2742.

Drozd P & Novotny V (2010) *AccuCurve. Version 1*. http://prf.osu.cz/kbe/dokumenty/sw/AccuCurve/AccuCurve.xls (Accessed January 13, 2013).

Eichler S, Christen R, Höltje C, Westphal P, Bötel J, Brettar I, Mehling A & Höfle MG (2006) Composition and Dynamics of Bacterial Communities of a Drinking Water Supply System as Assessed by RNA- and DNA-Based 16S rRNA Gene Fingerprinting. *Appl Environ Microbiol* 72: 1858–1872.

Fairs A, Wardlaw AJ, Thompson & Pashley CH (2010) Guidelines on ambient intramural airborne fungal spores. *J Investig Allergol Clin Immunol* 20: 490–498.

Fröhlich-Nowoisky J, Pickersgill DA, Després VR & Pöschl U (2009) High diversity of fungi in air particulate matter. *Proc Natl Acad Sci U S A* 106: 12814–12819.

Gast CJ van der, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J & Bruce KD (2011) Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME Journal* 5: 780–791.

Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A & Gillevet PM (2010) Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathog* 6: e1000713.

Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S & Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 5: 20–29.

Hauser AR, Jain M, Bar-Meir M & McColley SA (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clin Microbiol Rev* 24: 29–70.

Hazen KC, Wu JG & Masuoka J (2001) Comparison of the Hydrophobic Properties of Candida albicans and Candida dubliniensis. *Infect Immun* 69: 779–786.

Hibbett DS et al. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research* 111: 509–547.

Ifp Institut für Produktqualität (2008) *PCRFast® Mycobacterium tuberculosis for Human Diagnostics*. Berlin, Germany http://novatec-

id.com/pdf/pcr/englisch/PCRFast%20Mycobacterium%20tuberculosis_080924.pdf (Accessed January 8, 2013).

Jenkinson HF & Lamont RJ (2005) Oral microbial communities in sickness and in health. *Trends Microbiol.* 13: 589–595.

Kraneveld EA, Buijs MJ, Bonder MJ, Visser M, Keijser BJF, Crielaard W & Zaura E (2012) The relation between oral Candida load and bacterial microbiome profiles in Dutch older adults. *PLoS ONE* 7: e42770.

Kumar M & Shukla PK (2005) Use of PCR targeting of internal transcribed spacer regions and single-stranded conformation polymorphism analysis of sequence variation in different regions of rrna genes in fungi for rapid diagnosis of mycotic keratitis. *J. Clin. Microbiol.* 43: 662–668.

LiPuma JJ (2010) The Changing Microbial Epidemiology in Cystic Fibrosis. *Clin. Microbiol. Rev.* 23: 299–323.

Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW & Boucher RC (1998) Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 95: 1005–1015.

McAlester G, O'Gara F & Morrissey JP (2008) Signal-mediated interactions between Pseudomonas aeruginosa and Candida albicans. *J. Med. Microbiol* 57: 563–569.

Moran GP, Coleman DC & Sullivan DJ (2012) Candida albicans versus Candida dubliniensis: Why Is C. albicans More Pathogenic? *International Journal of Microbiology* 2012: 1–7.

Müller F-MC & Seidler M (2010) Characteristics of pathogenic fungi and antifungal therapy in cystic fibrosis. *Expert Rev Anti Infect Ther* 8: 957–964.

Mullins J & Seaton A (1978) Fungal spores in lung and sputum. *Clin. Allergy* 8: 525–533.

Muthig M, Hebestreit A, Ziegler U, Seidler M & Müller F-MC (2010) Persistence of Candida species in the respiratory tract of cystic fibrosis patients. *Med. Mycol.* 48: 56–63.

Parkins MD, Sibley CD, Surette MG & Rabin HR (2008) The Streptococcus milleri group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatr. Pulmonol.* 43: 490–497.

Peltroche-Llacsahuanga H, Döhmen H & Haase G (2002) Recovery of Candida dubliniensis from sputum of cystic fibrosis patients. *Mycoses* 45: 15–18.

Pihet M et al. (2009) Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis – a review. *Medical Mycology* 47: 387–397.

Ranjard L, Poly F, Lata J-C, Mougel C, Thioulouse J & Nazaret S (2001) Characterization of Bacterial and Fungal Soil Communities by Automated Ribosomal Intergenic Spacer Analysis Fingerprints: Biological and Methodological Variability. *Appl. Environ. Microbiol.* 67: 4479–4487.

Rantio-Lehtimäki A (1985) Mould spores and yeasts in outdoor air. *Allergy* 40 Suppl 3: 17–20.

Rogers GB, Carroll MP & Bruce KD (2009) Studying bacterial infections through culture-independent approaches. *J Med Microbiol* 58: 1401–1418.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G & Bruce KD (2004) Characterization of Bacterial Community Diversity in Cystic Fibrosis Lung Infections by Use of 16S Ribosomal DNA Terminal Restriction Fragment Length Polymorphism Profiling. *J. Clin. Microbiol.* 42: 5176–5183.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ & Bruce KD (2006) Use of 16S rRNA Gene Profiling by Terminal Restriction Fragment Length Polymorphism Analysis To Compare Bacterial Communities in Sputum and Mouthwash Samples from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 44: 2601–2604.

Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ & Bruce KD (2003) Bacterial Diversity in Cases of Lung Infection in Cystic Fibrosis Patients: 16S Ribosomal DNA (rDNA) Length Heterogeneity PCR and 16S rDNA Terminal Restriction Fragment Length Polymorphism Profiling. *J. Clin. Microbiol.* 41: 3548–3558.

Rudkjøbing VB, Thomsen TR, Alhede M, Kragh KN, Nielsen PH, Johansen UR, Givskov M, Høiby N & Bjarnsholt T (2011) True Microbiota Involved in Chronic Lung Infection of Cystic Fibrosis Patients Found by Culturing and 16S rRNA Gene Analysis ▽. *J Clin Microbiol* 49: 4352–4355.

Schoch CL et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* doi: 10.1073/ pnas.1117018109.

Sibley CD, Grinwis ME, Field TR, Eshaghurshan CS, Faria MM, Dowd SE, Parkins MD, Rabin HR & Surette MG (2011) Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS ONE* 6: e22702.

Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC & Surette MG (2008) A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci U S A* 105: 15070–15075.

De Sordi L & Mühlschlegel FA (2009) Quorum sensing and fungal–bacterial interactions in Candida albicans: a communicative network regulating microbial coexistence and virulence. *FEMS Yeast Research* 9: 990–999.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207–214.

Valenza G, Tappe D, Turnwald D, Frosch M, König C, Hebestreit H & Abele-Horn M (2008) Prevalence and antimicrobial susceptibility of microorganisms isolated from sputa of patients with cystic fibrosis. *J. Cyst. Fibros.* 7: 123–127.

Wos-Oxley ML, Plumeier I, Eiff C von, Taudien S, Platzer M, Vilchez-Vargas R, Becker K & Pieper DH (2010) A poke into the diversity and associations within human anterior nare microbial communities. *The ISME Journal* 4: 839–851.

Zaura E, Keijser BJF, Huse SM & Crielaard W (2009) Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol.* 9: 259.

Zhao J et al. (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci. U.S.A.* 109: 5809–5814.

3.8 Supplementary Material

S1: ***Bacteria***

OTU designation	taxonomic class	sputa no. pos. (%)	patients no. pos. (%)	RDP score	closest described bacterial species (% similarity, accession no.)	representative sequence (accession no., name)
Pseudomonas aeruginosa	Gammaproteo bacteria	50 (69.4 %)	38 (67.8 %)	1.000	100%_CP003149.1_Pseudomonas aeruginosa DK2	AB626118_Pseudomonas_aeruginosa strain__JCM_14847
Streptococcus parasanguis	Bacilli	41 (56.9 %)	36 (64.2 %)	0,990	99%_AY281078_Streptococcus mitis ATCC 903*	AF003933_Streptococcus_parasanguinis ATCC_15912
Streptococcus salivarius	Bacilli	41 (56.9 %)	35 (62.5 %)	1.000	100%_FR873482.1_Streptococcus salivarius JIM8777	AY188352_Streptococcus_salivarius strain_ATCC_7073
Streptococcus mitis	Bacilli	31 (43.0 %	29 (51.7 %)	0,988	99%_FN568063.1_Streptococcus mitis B6	DQ303188 Streptococcus mitis strain ATCC 49456
Rothia mucilaginosa	Actinobacteria	31 (43.0 %)	28 (50.0 %)	0,969	99%_DQ409140.1_Rothia mucilaginosa strain C90	X87758_Rothia_mucilaginosa DSM_20746
Staphylococcus aureus	Bacilli	18 (25.0 %)	15 (26.7 %)	1.000	99%_AB680391.1_Staphylococcus aureus subsp. aureus strain NBRC 13276	X68417_Staphylococcus_aureus ATCC_12600
Gemella sanguinis	Bacilli	16 (22.2 %)	14 (25.0 %)	1.000	100%_NR_026419.1_Gemella sanguinis strain 2045-94	Y13364_Gemella_sanguinis strain_2045-94
Granulicatella -2	Bacilli	14 (19.4 %)	13 (23.2 %)	1.000	100%_FR822389.1_Granulicatella adiacens strain CCUG 60768	D50540_Granulicatella_adiacens ATCC_49175
Actinomyces graevenitzii	Actinobacteria	12 (16.6 %)	11 (19.6 %)	0,964	99%_ NR_042167.1_Actinomyces graevenitzii strain CCUG 27294	AJ540309_Actinomyces_graevenitzii strain_CCUG_27294T
Gemella haemolysans	Bacilli	7 (9.7 %)	7 (12.5 %)	1.000	100%_NR_025903.1_Gemella haemolysans strain ATCC 10379	L14326_Gemella_haemolysans ATCC_10379
Streptococcus anginosus	Bacilli	6 (8.3 %)	5 (8.9 %)	0,986	99%_AF306838.1_Streptococcus anginosus genotype VA8466	AF169355 Streptococcus anginosus strain VAMC5219
Achromobacter xylosoxidans	Betaproteo bacteria	6 (8.3 %)	5 (8.9 %)	1.000	99%_EU266588.1_Achromobacter xylosoxidans strain CS5	Y14908_Achromobacter_xylosoxidans DSM_10346
Prevotella histicola	Bacteroidia	5 (6.9 %)	5 (8.9 %)	1.000	100%_AB547685.1_Prevotella histicola strain JCM 15637	AB547685_Prevotella_histicola strain_JCM_15637
Prevotella melaninogenica	Bacteroidia	5 (6.9 %)	5 (8.9 %)	0,985	99%_JN867309.1_Prevotella melaninogenica strain SEQ236	L16469_Prevotella_melaninogenica ATCC_25845
Fusobacterium nucleatum	Fusobacteria	5 (6.9 %)	4 (7.1 %)	0,951	98%_AB588016.1_Fusobacterium nucleatum subsp. nucleatum JCM 6328	AB573068_Fusobacterium_nucleatum subsp._nucleatum_strain__JCM_8532
Streptococcus intermedius	Bacilli	5 (6.9 %)	4 (7.1 %)	1.000	100%_AF104671.1_Streptococcus intermedius strain ATCC 27335	AF104671_Streptococcus_intermedius strain_ATCC27335
Veillonella -1	Negativicutes	4 (5.5 %)	4 (7.1 %)	0,985	99%_HM596287.1_Veillonella sp. oral taxon 158 strain F0412	X84007_Veillonella_sp. DSM_20739
Rothia -2	Actinobacteria	4 (5.5 %)	4 (7.1 %)	1.000	100%_AY594189.1_Rothia dentocariosa strain Inje L1	M59055_Rothia_dentocariosa ATCC_17931
Stenotrophomonas maltophilia	Gammaproteo bacteria	3 (4.1 %)	3 (5.3 %)	1.000	99%_JX848739.1_Stenotrophomonas maltophilia strain Sal10	AB294553_Stenotrophomonas maltophilia_strain_IAM_12423
Prevotella pallens	Bacteroidia	3 (4.1 %)	3 (5.3 %)	0,945	99%_AB547703.1_Prevotella pallens strain JCM 11140	Y13105_Prevotella_pallens strain__10371
Streptococcus gordonii	Bacilli	2 (2.7 %)	2 (3.5 %)	0,973	99%_AY281088.1_Streptococcus gordonii ATCC 12396	AF003931_Streptococcus_gordonii strain_ATCC_10558
Prevotella nanceiensis	Bacteroidia	2 (2.7 %)	2 (3.5 %)	0,976	99%_JN867319.1_Prevotella nanceiensis strain SEQ246	AB547695_Prevotella_nanceiensis JCM_15639
Prevotella oris	Bacteroidia	2 (2.7 %)	1 (1.7 %)	0,924	99%_AB547700.1_Prevotella oris strain JCM 12252	L16474_Prevotella_oris_ATCC_33573
Bordetella petrii	Betaproteo bacteria	1	1	1.000	100%_AJ870969.1_Bordetella petrii strain GDH030510	AJ249861_Bordetella_petrii strain_DSM_12804
Streptococcus sanguinis	Bacilli	1	1	1.000	100%_ AB596946.1_Streptococcus sanguinis JCM 5708	AF003928_Streptococcus_sanguinis ATCC_10556

Granulicatella elegans	Bacilli	1	1	1.000	100%_Y15413.1_Granulicatella elegans strain 4067-96	AF016390 Granulicatella elegans DSM 11693
Capnocytophaga gingivalis	Flavobacteria	1	1	0,942	99%_AB638448.1_Capnocytophaga gingivalis strain JCM 12953	AB638448_Capnocytophaga_gingivalis JCM_12953
Streptococcus australis	Bacilli	1	1	0,929	98%_NR_036936.1_Streptococcus australis strain ATCC 700641	AY485604 Streptococcus australis strain ATCC 700641
Actinomyces odontolyticus	Actinobacteria	1	1	1.000	100%_GQ131411.1_Actinomyces odontolyticus strain F0309	AJ234040_Actinomyces_odontolyticus CCUG_20536T
Burkholderia cepacia	Betaproteo bacteria	1	1	1.000	100%_AB680641.1_Burkholderia cepacia NBRC 14595	AF097530_Burkholderia_cepacia ATCC_25416
Prevotella pleuritidis	Bacteroidia	1	1	0,937	99%_ AB278593.1_Prevotella pleuritidis strain JCM 14110	AB278593_Prevotella_pleuritidis JCM_14110
Prevotella -8	Bacteroidia	1	1	0,741	90%_GU409582.1_Prevotella sp. oral taxon 306 clone DO004	GU409582_Prevotella_sp. oral_taxon_306
Prevotella tannerae	Bacteroidia	1	1	0,973	99%_AF183406.1_Prevotella tannerae strain 93-1-2	AJ005634_Prevotella_tannerae ATCC_51259
Nocardia sp.	Actinobacteria	1	1	1.000	100%_GU992878.1_Nocardia sp. W9912	GU992877_Nocardia_sp._W8474
Atopobium parvulum	Actinobacteria	1	1	0,985	99%_CP001721.1_Atopobium parvulum DSM 20469	AF292372_Atopobium_parvulum ATCC_22793
Streptococcus peroris	Bacilli	1	1	0,943	99%_GU425263.1_Streptococcus peroris clone GD031	AB008314_Streptococcus_peroris GTC_848
TM7	TM7 candidate class	1	1	0,766	99%_GU410609.1_TM7 phylum sp. oral taxon 352 clone DR034	AF385520_TM7_phylum_sp. oral_clone_DR034
Prevotella -9	Bacteroidia	1	1	0,670	99%_GU409603.1_Prevotella sp. oral taxon 308 clone GD054	GU409605 Prevotella sp. oral taxon 308 clone OCU043

Table S1: OTU designation for bacterial sequences that were found in sputa. For each OTU, accession numbers of closest described bacterial sequence are given, matching one representative sequence of ssDNA bands from corresponding clade in phylognetic distance tree (data not shown). Total number of positive (no. pos.) sputum samples and patients are shown and respective occurrence frequencies (%) for repeatedly detected OTUs are given in brackets. OTUs with ambiguous results in databases were defined according to the closest shared taxonomic level. Highest RDP scores for described species are shown. In representative sequence, accession numbers of sequences and species name used for phylogenetic trees in **Figure 6** are given.

*strain reclassified into *Streptococcus parasanguinis*

S2.a: ***Ascomycota, Saccharomycetes***

OTU designation	**taxonomic class**	**sputa** no. pos. (%)	**patients** no. pos. (%)	**closest described fungal species (% similarity, accession no., name)**	**representative sequence**
Candida albicans	Saccharomycetes	32 (44.4 %)	27 (48.2 %)	(26x) 100%_JX094781_Candida albicans strain ATCC 96901	JX094781
Candida dubliniensis	Saccharomycetes	17 (23.6 %)	14 (25.0 %)	(16x) 100%_DQ105856_ Candida dubliniensis strain DSM 13628	DQ105856
Saccharomyces cerevisiae	Saccharomycetes	14 (19.4 %)	13 (23.2 %)	(3x) 100%_AM900403_Saccharomyces cerevisiae strain MUCL28071	FN393997
Candida parapsilosis	Saccharomycetes	10 (13.8 %)	7 (12.5 %)	(9x) 100%_FJ872015_Candida parapsilosis strain ATCC 22019	FJ872015
Candida glabrata	Saccharomycetes	9 (12.5 %)	6 (10.7 %)	100%_AF336836_Candida glabrata strain ATCC 2001	AF336836
Candida tropicalis	Saccharomycetes	3 (4.1 %)	3 (5.3 %)	100%_AY939810_Candida tropicalis strain ATCC 750	AY939810
Cyberlindnera jadinii (synonyms: Candida utilis, Pichia jadnii)	Saccharomycetes	3 (4.1 %)	3 (5.3 %)	(2x) 100%_FJ865435_Pichia jadinii isolate M9	99%_DQ249199_ Cyberlindnera jadinii strain WM 45
Candida sake	Saccharomycetes	2 (2.7 %)	2 (3.5 %)	98%_AJ549822_Candida sake strain CBS 159	AJ549822
Galactomyces geotrichum	Saccharomycetes	1	1	100%_EU789402_Galactomyces geotrichum strain M0163	EU789402
Meyerozyma guilliermondii (synonyms: Pichia guilliermondii)	Saccharomycetes	1	1	100%_JQ425356_Meyerozyma guilliermondii strain AUMC 7771	JQ425356
Kazachstania unispora *(synonym: Saccharomyces unisporus)*	Saccharomycetes	1	1	100%_EU789404_Kazachstania unispora strain M01621	EU789404
Candida humilis	Saccharomycetes	1	1	100%_AY188851_Candida humilis strain CBS 6897	AY188851
Debaryomyces hansenii	Saccharomycetes	1	1	100%_HE967326_Debaryomyces hansenii strain MRL2	HE967326
Clavispora lusitaniae	Saccharomycetes	1	1	100%_JN391310_Clavispora lusitaniae isolate PUMY044	100%_EF568023_ Clavispora lusitaniae strain WM 1138
Yarrowia lipolytica	Saccharomycetes	1	1	100%_ EU252546_Yarrowia lipolytica ATCC 9773	EU252546
Kluyveromyces marxianus	Saccharomycetes	1	1	100%_HQ014731_Kluyveromyces marxianus strain WM10.112	HQ014731
Candida deformans	Saccharomycetes	1	1	100%_FJ515168_Candida deformans strain SM21	FJ515168
Candida cellae	Saccharomycetes	1	1	100%_GQ149495_Candida cellae strain UAF-93	GQ149495
Hanseniaspora uvarum	Saccharomycetes	1	1	100%_AJ512432_Hanseniaspora uvarum strain CBS 314	AJ512432

S2.b: ***Ascomycota,*** **non-*Saccharomycetes***

OTU designation	taxonomic class	sputa no. pos. (%)	patients no. pos. (%)	closest described fungal species (% similarity, accession no., name)	representative sequence
Cladosporium cladosporioides (teleomorph: Davidiella sp.)	Dothideomycetes	8 (11.1 %)	7 (12.5 %)	(5x) 100%_JQ768323_Cladosporium cladosporioides strain CFP14	JQ768323
Cladosporium herbarum (teleomorph: Davidiella tassiana)	Dothideomycetes	7 (9.7 %)	6 (10.7 %)	(5x) 100%_HQ263359_Davidiella tassiana strain ATCC MYA-4682	HQ263359
Scedosporium apiospermum (teleomorph: Pseudallescheria boydii)	Sordariomycetes	3 (4.1 %)	2 (3.5 %)	(2x) 100%_AY213682_Pseudallescheria boydii strain UWFP 806	AY213682
Exophiala dermatitidis	Eurotiomycetes	3 (4.1 %)	2 (3.5 %)	100%_JX473286_Exophiala dermatitidis strain PW2643	JX473286
Aspergillus fumigatus	Eurotiomycetes	2 (2.7 %)	2 (3.5 %)	99%_HQ026746_Aspergillus fumigatus strain ATCC 1022	HQ026746
Blumeria graminis	Leotiomycetes	2 (2.7 %)	2 (3.5 %)	100%_AB273556_Blumeria graminis isolate MUMH2335	AB273556
Lewia infectoria (anamorph: Alternaria infectoria)	Dothideomycetes	2 (2.7 %)	2 (3.5 %)	100%_FJ214897_Lewia infectoria strain CBS 112250	FJ214897
Coniothyrium fuckelii (teleomorph: Leptosphaeria coniothyrium)	Dothideomycetes	1	1	98%_FJ228185_Coniothyrium fuckelii isolate 7a-1	98%_FJ861383_ Coniothyrium fuckelii strain 3T16A
Aureobasidium pullulans	Dothideomycetes	1	1	100%_AF121281_Aureobasidium pullulans strain ATCC11942	AF121281
Phaeosphaeria sp.	Dothideomycetes	1	1	98%_AF439488_Phaeosphaeria juncophila isolate CBS 575.86	HM172819
Leotiomycetes -2	Leotiomycetes	1	1	97%_JN995647_Phialocephala sp. AU_BD20	JX317257
Ascomycota -22	unknown	1	1	89%_FJ487945_Pichia norvegensis strain ZH2Aea	87%_EU343827_ Candida ernobii strain MUCL 29961
Aspergillus conicus	Eurotiomycetes	1	1	99%_EF652039_Aspergillus conicus isolate NRRL 149	EF652039
Neosartorya pseudofischeri (anamorph: Aspergillus thermomutatus)	Eurotiomycetes	1	1	100%_AF459729_Neosartorya pseudofischeri NRRL 180	AF459729
Leptosphaeria sp.	Dothideomycetes	1	1	99%_EU852362_Uncultured Leptosphaeria clone 21a	EU852362
Didymella exitialis	Dothideomycetes	1	1	100%_EU167564_Didymella exitialis strain CBS 446.82	EU167564
Phoma exigua	Dothideomycetes	1	1	100%_EU770244_Phoma exigua strain ICMP 16989	EU770244
Aspergillus versicolor	Eurotiomycetes	1	1	100%_EF652449_Aspergillus versicolor isolate NRRL 239	EF652449
Septoria sp.	Dothideomycetes	1	1	100%_JX480493_Septoria erigerontis voucher KUS-F25759	JX480493
Penicillium sp.	Eurotiomycetes	1	1	100%_JN983439_Penicillium sp. mcp2728	JN983439
Fusarium oxysporum	Sordariomycetes	1	1	100%_HQ829111_Fusarium oxysporum strain CID 207	HQ829111
Colletotrichum circinans	Sordariomycetes	1	1	99%_EU400140_Colletotrichum circinans strain DAOM151616	EU400140

S2.c: *Basidiomycota*

OTU designation	taxonomic class	sputa no. pos. (%)	patients no. pos. (%)	closest described fungal species (% similarity, accession no., name)	representative sequence
Sporobolomyces roseus	Puccinio mycetes	8 (11.1 %)	7 (12.5 %)	(7x) 100%_AY069997_Sporobolomyces roseus strain CBS 993	AY069997
Sporobolomyces ruberrimus	Puccinio mycetes	4 (5.5 %)	3 (5.3 %)	(3x) 100%_JN246564_Sporobolomyces ruberrimus strain CRUB 1041	JN246564
Rhodotorula glutinis	Microbotryo mycetes	2 (2.7 %)	2 (3.5 %)	100%_AF335948_Rhodotorula glutinis strain ATCC 32765	AF335948
Udeniomyces pannonicus	Tremello mycetes	1	1	100%_AB072231_Udeniomyces pannonicus strain JCM 11148	AB072231
Plicaturopsis crispa	Agarico mycetes	1	1	100%_DQ534576_Plicaturopsis crispa strain FP-101310-SP	100%_HQ871872_ Plicaturopsis crispa voucher TENN53066
Rhodotorula -2	Microbotryo mycetes	1	1	100%_AM160641_Rhodotorula sp. HB 1211	99%_AB026015_ Rhodotorula aurantiaca strain JCM 3771
Basidiomycota -18	unknown	1	1	77%_DQ411529_Trechispora alnicola isolate AFTOL-ID 665	80%_FR682413_ Uncultured Basidiomycota clone BF-OTU670
Megacollybia platyphylla	Agarico mycetes	1	1	100%_EU623713_Megacollybia platyphylla voucher LE 256-2004	EU623713
Cryptococcus -3	Tremello mycetes	1	1	96%_JN400817_Cryptococcus tephrensis	99%_AM160648_ Cryptococcus sp. HB 1222
Cryptococcus victoriae	Tremello mycetes	1	1	100%_AM160647_Cryptococcus victoriae strain HB 1221	AM160647
Agaricomycetes -8	Agarico mycetes	1	1	99%_GQ221186_Undifferentiated Agaricomycetes strain FSU6258	GQ221186
Malassezia restricta	Exobasidio mycetes	1	1	100%_JQ088240_Malassezia restricta isolate HIV+92	JQ088240
Piptoporus betulinus	Agarico mycetes	1	1	100%_DQ491423_ Piptoporus betulinus strain CBS 378.51	JQ700297
Polyporus gayanus	Agarico mycetes	1	1	100%_AF518757_ Polyporus gayanus strain CIEFAP136	AF518757
Strobilurus sp.	Agarico mycetes	1	1	100%_GQ892818_Strobilurus albipilatus voucher TFB11910	GQ892818
Baeospora sp. *	Agarico mycetes	1	1	95%_EU770252_Baeospora sp. ICMP 16979	EU770252
Hyphodontia sp.	Agarico mycetes	1	1	100%_FJ197950_Uncultured Hyphodontia clone Bo2gLR21_A21	FJ197950
Cryptococcus curvatus	Tremello mycetes	1	1	100%_EU266558_Cryptococcus curvatus strain ATCC 10567	EU266558

Table S2: OTU designation for fungal sequences that were found in sputa, grouped by **a:** *Saccharomycetes*; **b:** *Ascomycota* (non-*Saccharomycetes*); **c:** *Basidiomycota*. Names of teleomorphs have priority. The anamorph name is only given if asexual state is reported more frequently from clinical samples. Common synonyms of species are given in brackets. Total number of positive (no. pos.) sputum samples and patients are shown and respective occurrence frequencies (%) for repeatedly detected OTUs are given in brackets. For each OTU, accession numbers of closest described fungus matching the most sequences of ssDNA bands are mentioned (number shown in brackets). In representative sequence, accession numbers of sequences used for phylogenetic trees (**Figure 8 & 9**) are shown and matching the sequences given for closest described fungus, unless a similarity score and accession number are additionally mentioned.

* ITS exhibit 100 % sequence similarity with *Mucronella sp.* PDD 95742

S3: Species accumulation curves

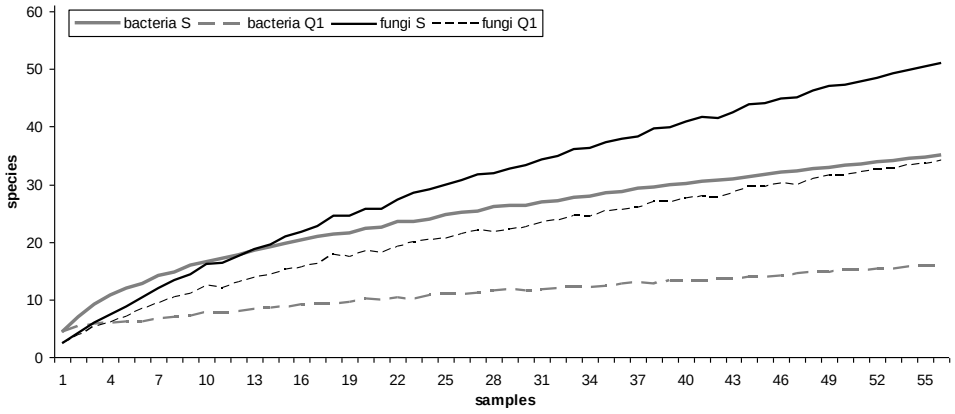


Figure S3: Species accumulation curves are shown for bacterial and fungal OTUs detected for 56 CF patients. Only the first sample of each patient was considered. Averaged curves show the number of detected species versus number of samples. Curves for bacteria are shown in grey and curves for fungi are shown in black. Full lines show observed species counts (S; Mao Tao) and dashed lines show unique species counts (species which occur in only one sample).

Chapter IV

Monitoring the VOCs composition of major microbial pathogens for clinical diagnostic purposes

Rolf Kramer[1], Annette Sauer-Heilborn[2], Carlos A. Guzman[1],
Manfred G. Höfle[1] and Wolf-Rainer Abraham[1]

[1]Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2]Department of Pneumology, Hannover Medical School (MHH), Germany

4.1 Abstract

Non-invasive diagnostic strategies focusing on the detection of volatile biomarkers in exhaled breath are of increasing interest. Volatile organic compounds (VOCs) have the potential to be such diagnostic markers. VOCs are released by all organisms, including infectious microorganisms colonizing the respiratory tracts of infected hosts. In the current study, a breath analytical method was developed to discriminate major pulmonary pathogens associated with cystic fibrosis (CF) and the feasibility of the method was tested in a CF outpatient clinic *in vivo*. VOCs were extracted by a sorbent trap and analyzed by gas chromatography and mass spectrometry. Specific microbial VOC compositions were used to discriminate the bacterial species *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Achromobacter xylosoxidans* and the yeast *Candida albicans in vitro*. Exhaled breath of CF patients was analyzed and allowed distinction of CF patients and controls by the individual *in vivo* VOC compositions. Although none of the specific compounds monitored *in vitro* were detected in exhaled breath of infected CF patients *in vivo*, the current study support the hypothesis that the composition of exhaled VOCs in human breath may be used to identify patients with certain microbial infections.

4.2 Introduction

Bacteria and fungi are known to produce a broad spectrum of secondary metabolites including a wealth of volatile molecules (Schulz & Dickschat 2007) (Brakhage & Schroeckh 2011). A large number of these are volatile organic compounds (VOCs) comprising aliphatic and aromatic hydrocarbons, esters, ketones, aldehydes, alcohols and mono-, sesqui- and diterpenes. Synthesis of VOCs can represent an advantage for the producing organism. Ecological functions have been demonstrated in fungi where VOCs defend against competing species, act as autoinducer or play an important role in quorum sensing (Kramer & Abraham 2012). Similar biological effects are also known in bacteria, where VOCs are produced under aerobic as well as anaerobic conditions (Kai et al. 2009) (O'Brien & Wright 2011). Because of these essential functions, VOCs can be considered to be almost ubiquitously released by microorganisms. However, the exact composition of released metabolites and, therefore, also of the emitted volatile compounds, varies with different substrates. Most of the microbial VOCs are released in very low concentrations but some are familiar to us as special odours of food products or even bacterial strains. In the past years, modern analytics enable the detection and monitoring of VOCs in the environment even at low concentrations (Woolfenden 2010b) (Woolfenden 2010a). Consequently, there is an increasing interest in the search for volatile microbial biomarkers and first databases for volatile profiles of microorganism have been established, e.g. "SuperScent" (Dunkel et al. 2009).

But not only in environmental or food-related research microbial VOCs are shifting in the focus, their potential for medical purposes have been recognized as well (Thomas et al. 2010) (Shirasu & Touhara 2011). Already in the 1970s, the use of information on volatile compounds for clinical diagnosis was investigated (Liddell 1976) (Zlatkis et al. 1981). First techniques for detection of volatiles in exhaled breath of humans were developed in order to monitor altered emission of human metabolites from malfunction of organs and tissues (Teranishi et al. 1972) (Zlatkis et al. 1973). A diagnostic potential also lies in the smell of certain VOCs at high concentrations. If released by inflammation or microbial colonization in the mouth they can cause an oral malodour. On that basis the University of the West of England started courses for dentists, training their 'nose' to differentiate the origins of the malodour and then treat appropriately (UWE Bristol 2012). Other technical developments intend to extract microbial VOCs from the exhaled air of patients suffering from respiratory tract infection. The aim is to detect specific infections by species-specific compounds that are released in concentrations high enough to be monitored in larger air volumes like the exhaled

breath. Active pulmonary tuberculosis was one of the first infectious diseases screened for volatile biomarkers in human exhalation (Phillips et al. 2007) (Syhre & Chambers 2008) (Syhre et al. 2009) (Phillips et al. 2010). However, while these studies could greatly show the potential for a breath analytical diagnostic tool in tuberculosis, they did not fulfil the needs for routine diagnostics or revealed any reliable marker compound. Another promising microorganism in this context is the lung infecting fungus *Aspergillus fumigatus*. A single VOC with high potential to act as biomarker was identified and, being not part of the human metabolism, detected in the exhaled breath of infected patients but not in healthy individuals (Syhre et al. 2008) (Chambers et al. 2009) (Chambers et al. 2011). Likewise, a potential breath biomarker was investigated for the diagnosis of *Pseudomonas aeruginosa* infections in cystic fibrosis (CF) patients. 2-Aminoacetophenone has been verified *in vitro* for its specificity and exhaled breath of CF patients was screened for increased levels. Although infected patients showed higher levels, the same VOC was also present in breath samples of healthy and non-colonized individuals, leaving doubts about its reliability (Scott-Thomas et al. 2010a). Nevertheless, particularly in the context of the recessive genetic disorder CF, analytical diagnostics for infectious pulmonary diseases using VOCs could be of greater interest. CF patients suffer from live-long recurrent microbial colonization of their respiratory tracts and pulmonary infections are the main cause of death (Harrison 2007). A vast spectrum of pathogens can infect the airways and therefore a rapid, non-invasive diagnostic tool like exhaled air analysis would substantially improve the health care of these patients. Additionally, lung function tests are already part of routine medical examinations, presumably leading to a greater acceptance among patients for an additional breath diagnostic procedure.

The purpose of this study was to validate the discriminative potential of VOCs emitted by major respiratory pathogens *in vitro* and of exhaled VOCs likewise extracted from breath of CF patients *in vivo*. The obtained results suggest that VOCs may be used to identify different microorganisms. Furthermore, individual VOCs compositions monitored in exhaled breath allowed distinction between CF patients and controls.

4.3 Material and Methods

4.3.1 Cell culture and microbial strain preparation

Precultures of microorganisms were grown overnight in Luria Broth (LB) liquid medium and diluted with fresh LB medium. The following strains were allowed to grow until the mid-log phase: *Staphylococcus aureus* strain DSM 20231, *P. aeruginosa* strain PaO1, *Achromobacter xylosoxidans* strain DSM 2402, and *Candida albicans* strain SC 5314. Cultures were allowed to grow until the mid-log phase. All liquid LB cultures were incubated at 37°C on a shaking platform at 120 rpm. Microbial suspensions were used to determine colony-forming units (CFU) and *in vitro* VOCs extractions. For determination of CFUs, suspensions were diluted to appropriate concentrations and triplicates of 200 µl each were plated on LB agar plates. Agar plates were incubated at 37°C overnight or 3 days for *A. xylosoxidans* before colonies were counted.

The A549 human epithelial cell line was maintained in Dulbecco's modified Eagle's medium (DMEM; PAA, Pasching, Austria), with 10% foetal bovine serum (FCS) and 2 mM L-glutamine without antibiotics. Cells were incubated in culture flasks with growth area of 25 cm^2 at 37°C in a humidified atmosphere with 5% CO_2. Before reaching confluency, cells were trypsinized and sub-cultured.

4.3.2 Microbial *in vitro* models for measurements of VOC emissions

Biofilm models were defined as monocultures in agar slants using precultures of microorganisms grown overnight in LB liquid medium. 100 µl of the cultures in stationary phase were taken and plated on angular LB agar in a 15 ml glass vial. Vials were closed with rubber septa and screw caps. Incubation was done in 37°C overnight, before a SPME fiber was exposed through the septum to the headspace of the angular LB agar.

Host-pathogen interaction cultures were defined as A549 cell cultures inoculated with microbial pathogens. Therefore, 1 ml of the bacterial suspensions and 2 ml of the fungal suspensions in the mid-log phase were taken and centrifuged at 10,000 rpm for 8 min. Supernatant was discarded and the cell pellet resuspended in the same volume using DMEM (PAA, Pasching, Austria) with 10% FCS and 2 mM L-glutamine without antibiotics. 200 µl and 2 ml of bacterial and fungal suspensions, respectively, were added to a confluent monolayer of A549 cells and a SPME fiber was exposed to the headspace of the culture at different time points during the course of interaction (**Table 1**). For overnight incubation, bacterial suspensions were further diluted in the same medium by the following factors:

$1x10^{-2}$ for *P. aeruginosa*, $1x10^{-2}$ for *S. aureus* and $1x10^{-1}$ for *A. xylosoxidans*. 200 µl each was given to a confluent monolayer of A549 cells and SPME fiber exposed after 21 to 22.5 hours of incubation. For *C. albicans*, 200 µl of the undiluted suspension were used for overnight incubation and the SPME fiber was exposed after 20.5 hours. Initial pathogen inoculum added to A549 cells was $7.6x10^8$ ($7.6x10^7$) CFUs for *A. xylosoxidans*, $4.8x10^8$ ($4.8x10^6$) CFUs for *S. aureus*, $2.7\ x10^7$ ($2.7x10^6$) CFUs for *C. albicans*, and $9.5x10^8$ ($9.5x10^6$) CFUs for *P. aeruginosa* (numbers in brackets indicate CFUs injected to overnight incubation cultures). All cells were kept at 37°C in a humidified atmosphere with 5% CO_2.

Corresponding monocultures inoculated with pathogens in their mid-log phase, like it was done for the host-pathogen interaction cultures, were defined as monocultures in agar slants using 100 µl of microbial suspensions (LB medium) and plated on angular LB agar in 15 ml glass vials. Vials were closed with rubber septa and screw caps. Incubation was done in 37°C overnight, before a SPME fiber was exposed through the septum to the headspace of the angular LB agar.

C. albicans						
T0	T1	T2	T3	T4	T5	Tmax
0 min	95 min	195 min	275 min	390 min	470 min	20.5 h
P. aeruginosa						
T0	T1	T2	T3	T4	T5	Tmax
0 min	115 min	215 min	295 min	410 min	520 min	22.5 h
A. xylosoxidans						
T0	T1	T2	T3	T4		Tmax
0 min	120 min	210 min	300 min	390 min	no extraction	21 h
S. aureus						
T0	T1	T2	T3	T4		Tmax
0 min	120 min	210 min	300 min	390 min	no extraction	21 h

Table 1: Time points of SPME fiber exposition in host-pathogen interaction cultures. A series of VOCs extractions (T0 - Tmax) were performed at different time points post injection (min = minutes; h = hours) for each culture. The pathogen of each interaction culture is given in italic letters above corresponding rows.

4.3.3 Imaging of host-pathogen interaction cultures

To monitor growth of pathogens, cultures were visually controlled and digitally documented using Axiovert 200M inverted microscope coupled with AxioCam HRc (Carl Zeiss, Göttingen, Germany). Subsequently to SPME fiber exposition, images were taken with 10x and 20x objective lenses. To monitor shape and confluence state of A549 cells, cultures were washed three times with PBS after VOCs extraction T4 or T5 respectively. Laser scanning microscopy (LSM) was performed using the Leica SP5 system (Leica, Heidelberg, Germany).

Interaction cultures with *C. albicans* were stained with green fluorescent nucleic acid stain SYTO 13 and chitin binding fluorescent stain Calcofluor-white. Cultures were scanned in bidirectional mode using a 63 x 0.9 NA objective lens, laser diode (405 nm) and argon laser (488 nm) with scanner frequency of 400 Hz and line averaging of 4.

4.3.4 Solid phase micro-extraction (SPME)

VOCs were extracted using solid phase microextraction (SPME), where the volatiles are trapped on a sorbent coated fiber. Various VOCs extractions were performed *in vitro* and in *vivo* with healthy individuals and CF patients to select the proper fiber coating for this particular application. In addition to the four major criteria for the optimal fiber, (i) molecular weight and size of the analytes (ii) polarity (iii) concentration level and (iv) complexity of sample (Pawliszyn 2009), also the storage and transport between the clinic (Hannover Medical School) and the institute (Helmholtz Centre for Infection Research) had to be considered. In order to keep the experimental design as comparable as possible, the same fiber coatings were used for *in vitro* and *in vivo* applications. These demands were best fulfilled by a 75 µm Carboxen/PDMS fiber (Portable Field Sampler) from Supelco (Bellefonte, PA, USA). All fibers were preconditioned for 20 min in the injection port at 250°C before using or being brought to the clinic. The sorbent coated fiber was exposed for 10 min to the sample and, subsequently, VOCs were released by thermal desorption in the injection port of the GC at 200°C. Even after preconditioning, an individual peak pattern ubiquitously present in all associated chromatograms was observed for each SPME fiber. To minimize bias in the data analysis caused by these individual peak patterns, all fibers used for *in vivo* applications were numbered and indicated for each sample. For *in vitro* experiments the same fiber was taken for a series of associated VOCs extraction if possible.

4.3.5 Gas chromatography/mass spectroscopy (GC/MS) parameters

A Varian 450-GC gas chromatograph coupled with ion-trap mass spectrometer Varian 240-MS (Varian, Palo Alto, USA) was used to perform GC/MS analysis. Injection of SPME fiber was done manually at 200°C and with splitless mode for 2 min, followed by a split ratio of 100 for another 2 min and a final split ratio of 20. The column was a forte GC Capillary Column 50 m x 0.32 mm x 1 µm (SGE Analytical Science, Ringwood, Australia). Temperatures of ion trap, manifold and transfer line were 150°C, 50°C and 180°C respectively. The GC oven temperature was programmed from 35°C (hold for 2 min) to 250°C in two steps with a total analysis time of 32 min: first the column was heated to 75°C

at 3°C min^{-1}, followed by 15°C min^{-1} until final temperature was reached (hold for 5 min). Helium flow was set at constant rate of 1 ml min^{-1}. Analyses were done in EI full scan mode with a mass range from 40 *m/z* to 205 *m/z*.

4.3.6 CF cohort

CF patients were recruited in the CF outpatient clinic of the Hannover Medical School (MHH; Hannover, Germany), after ethical approval for this study was granted by the local health authority ethics committee. All patients were between 30 - 48 years old and of both genders. Breath samples were taken during their routine examination in the ambulance. Anonymized clinical data of patients was collected, including culture-based microbiological diagnostics (**Table S1**). Additionally, breath samples of two healthy probands (30 years old; male and female) were taken in the Helmholtz Centre for Infection Research (Braunschweig, Germany).

4.3.7 *In vivo* breath sampling

A commercially available device was used for the breath sampling (BIO-VOC® sampler, Markes International Ltd, Rhonda Cynon Taff, UK) (Dyne et al. 1997)(Henderson & Matthews 2002). After carrying out a series of exhaled air samplings with healthy probands and CF patients to establish a comparable and reliable sampling method, both the equipment and the procedure recommended by the manufacturer was modified. Self-made autoclavable mouthpieces and an adequate adapter were designed (made of polyvinyl chloride) and adjusted to the breath sampler, in order to guarantee a sterile procedure and easy handling for the patients and the medical staff. Patients and healthy volunteers were asked to rest and, when they feel relaxed, to exhale out completely a single slow breath into the sampler. The last portion of the exhaled air, originating from the alveolar space, was trapped (approx. 100 ml) and a SPME fiber was immediately exposed to the sample as described above. No cross-contamination of exhaled VOCs was observed when breath sampler was kept unclosed for two hours to release the remaining analytes. Additional room samples were taken to measure VOCs in the ambient air of the sampling location. For this purpose, SPME fiber was exposed to the empty sampler without human breath.

4.3.8 Data processing and statistical analysis

GC-MS data was analyzed using the MS Data Review (Version 6.9.3) from Varian. Since injection was done manually, all peaks had to be compared by retention times and mass

spectra to enable alignment of chromatograms. Parameters for peak integration of *in vitro* data were the following: Peak Width. 4.0 sec, Slope Sensitivity 20 SN, Tangent 30%. For the host-pathogen interaction cultures a minimum peak size of 10,000 units was chosen for most microorganisms and 5,000 units for *P. aeruginosa* respectively. For VOCs extractions from biofilms, no specific minimum peak size was defined. Parameters for peak integration of the *in vivo* breath sampling data were: Peak Width. 4.0 sec, Slope Sensitivity 2 SN, Tangent 10% and a minimum peak size of 10,000 units. After alignment of integrals from the different chromatograms, several peaks were rejected to exclude non-informative data points. For host-pathogen interaction cultures, peaks were rejected that were not detectable in corresponding monocultures on LB agar and appeared additionally (i) just once, (ii) just twice but not in the last time points and without an increase of size, or (iii) just thrice but not in consecutive time points and not of increasing size. Similarly, peaks were excluded for the *in vivo* exhaled breath data: peaks that appeared in several controls and just with smaller peak size in breath samples, peaks detected in controls and just in breath samples of healthy individuals, as well as peaks present in all samples without a significant change of peak size. For further analysis all data were divided by minimum peak size (5,000 units for *in vitro* and 10,000 units for *in vivo* data) and overall transformed (individual factors are mentioned). A maximum peak size before transformation of 100 was defined for the *in vivo* exhaled breath data. Heatmaps were calculated using the software package *R* (R Development Core Team 2011) from the colour palettes, the topo.colors package was chosen without scaling to either rows or columns.

Further statistical analyses were carried out using PRIMER (Version.6.1.6, PRIMER-E, Plymouth Marine Laboratory, UK). Resemblance matrices were calculated using Euclidean distance algorithm, taken all variables of a profile into account and providing the shortest statistical distance between every pair of samples. Euclidean distance is the square root of the sum of squared differences between profiles, the data of these has to be measured on the same scale. The samples were ordinated in non-metric multidimensional scaling (MDS) plots (with 50 random restarts), distances between samples in these plots reflect differences in the VOC compositions monitored for the samples. Trajectories were used to connect associated data points in chronological order. For each MDS plot, 2D stress values indicate the differences between ideal distances in higher dimensional space and the actual distances in two dimensional space. Therefore, 2D stress values are a measure for the reliability of the MDS plots. Additional analysis for *in vivo* data was performed using BEST analysis with BIOENV algorithm, which tries all variable combinations to find the ones “best explaining” the ordination pattern of the samples.

4.4 Results

Volatile organic compounds were monitored in different experimental set-ups, which are presented in consecutive order. Further identification of emitted compounds was not considered to be necessary for this diagnostic manner. Retention times and mass spectra revealed sufficient assessment for peak differentiation.

4.4.1 *In vitro* VOCs monitoring in biofilms of CF pathogens

Fig. 1A:

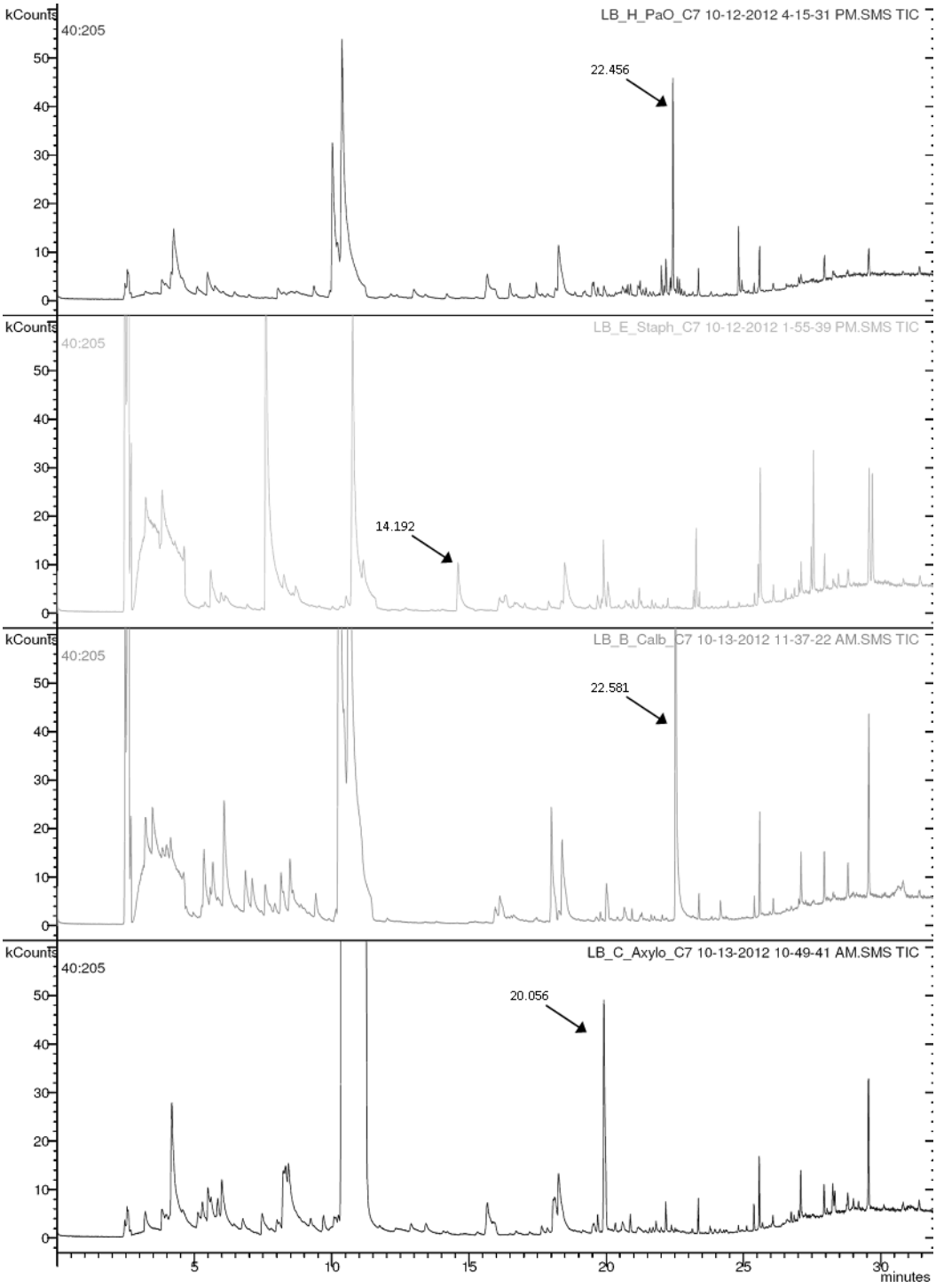


Fig. 1B:

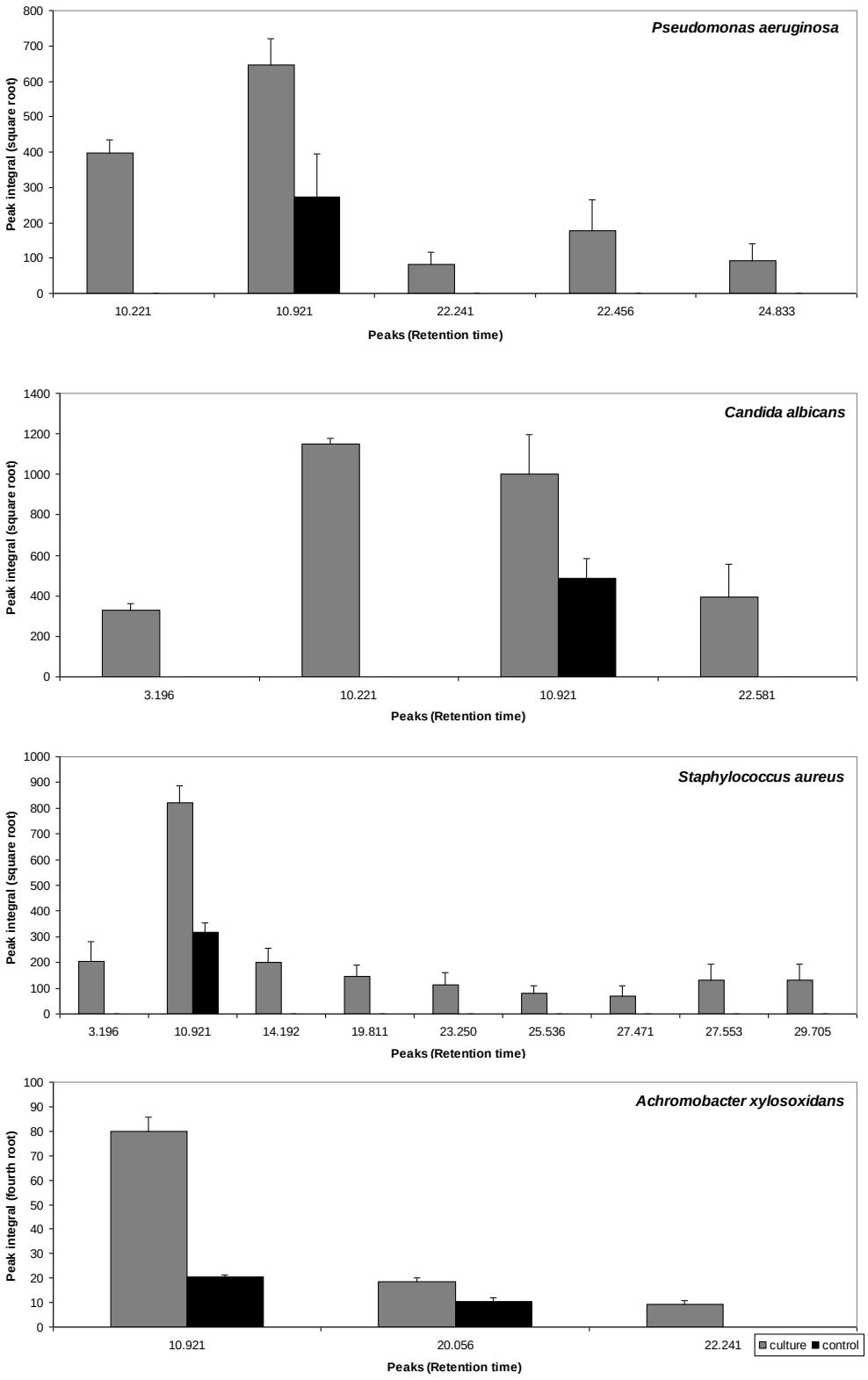


Figure 1A, B: VOCs detected from pathogens grown in biofilms on LB agar. **A**. Chromatograms show representative peak profiles in the following order (from top to bottom): *P. aeruginosa*, *S. aureus*, *C. albicans* and *A. xylosoxidans*. Arrows indicate conspicuous individual peaks highlighted for better orientation. **B.** Diagrams show integrals from conspicuous individual peaks released in biofilms. Overall transformation factors are mentioned in each diagram. Retention times of peaks are given on x-axes. Standard deviations are shown by error bars.

VOCs were extracted from microbes of relevance for CF patients, grown in biofilms on LB agar. The resulting chromatograms are shown in **Figure 1A**. After overnight incubation the cultures revealed different peak profiles for each pathogen. In addition, VOCs extractions were done as a control before the cultures had been started and used for comparison to identify unique or remarkable peaks for each microorganism. By this conspicuous individual VOCs were characterized for each biofilm. Just peaks ubiquitously present in all chromatograms of one pathogen but not in the controls were considered. Individual intensities did vary in different extractions and means of integrals are shown in **Figure 1B**. The following conspicuous VOCs were assessed: Peak at 10.921 min was present in all extractions and although also detectable before cultures were started, its intensities increase considerably with the presence of a microorganism. For *P. aeruginosa*, in total five conspicuous individual VOCs were detected. Only peaks at 22.456 min and 24.833 min were unique to the biofilm of this microorganism the others were likewise detected for the different pathogens. Four VOCs were found for *C. albicans*, from which just the one at 22.581 min was unique to the fungus. The peak at 10.221 min was also seen for *P. aeruginosa* and the one at 3.196 min also for *S. aureus* cultures in which extractions revealed in total nine, mostly unique peaks. *A. xylosoxidans* showed the smallest number of conspicuous VOCs. Just one of the four was unique to this betaproteobacterium. VOC at 20.056 min was also detectable in the controls but with much less intensities. A remarkable increase was observed for VOC at 10.921 min especially in comparison to the other microorganisms: in this biofilm, the peak integral (square root transformation) is 6,408 whereas the others show peak areas between 645 and 1,000. Additionally, VOC at 22.241 min was also present in cultures with *P. aeruginosa*. In summary, comparisons of profiles from monoculture biofilms reveal an individual VOC composition for each pathogen in culture, allowing discrimination among the different biofilms.

4.4.2 *In vitro* VOCs monitoring in host-pathogen interaction cultures (A549 cells)

Fig. 2A:

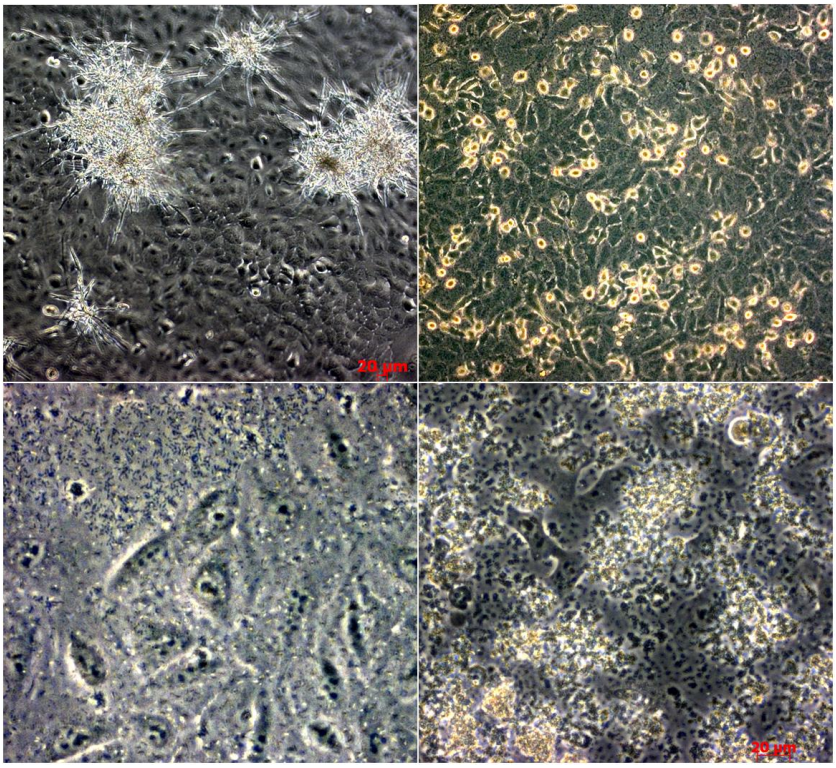


Fig. 2B:

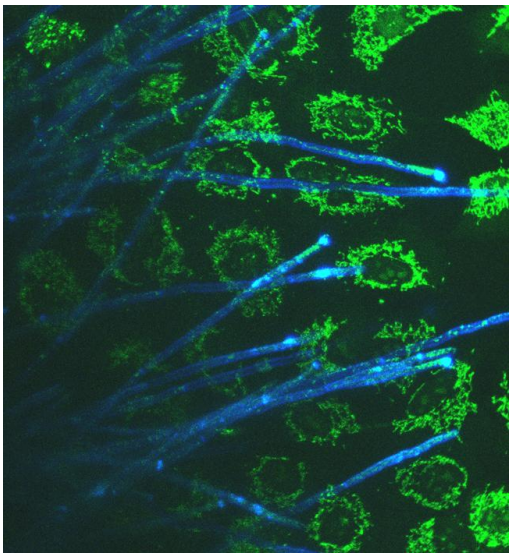


Figure 2A, B: Imaging of host pathogen interaction cultures. **A.** Light microscopy was performed during each series of VOCs extractions. Upper pictures: on the left, *C. albicans* was growing in hyphal form on a confluent monolayer of A549 cells. On the right, *P. aeruginosa* interaction culture, where the cell monolayer showed increasing signs of detachment during incubation. Both pictures are done with 10x objective lens. Lower picture: on the left, *A. xylosoxidans* in co-culture with apparently confluent A549 cells. On the right, *S. aureus* was growing in typical clusters in the cell culture medium. A 20x objective lens was used for these images.
B. LSM imaging for *C. albicans* on A549 cells. Fungal hyphae appeared in blue and human cells were fluorescenting in green.

A. Peak profiles

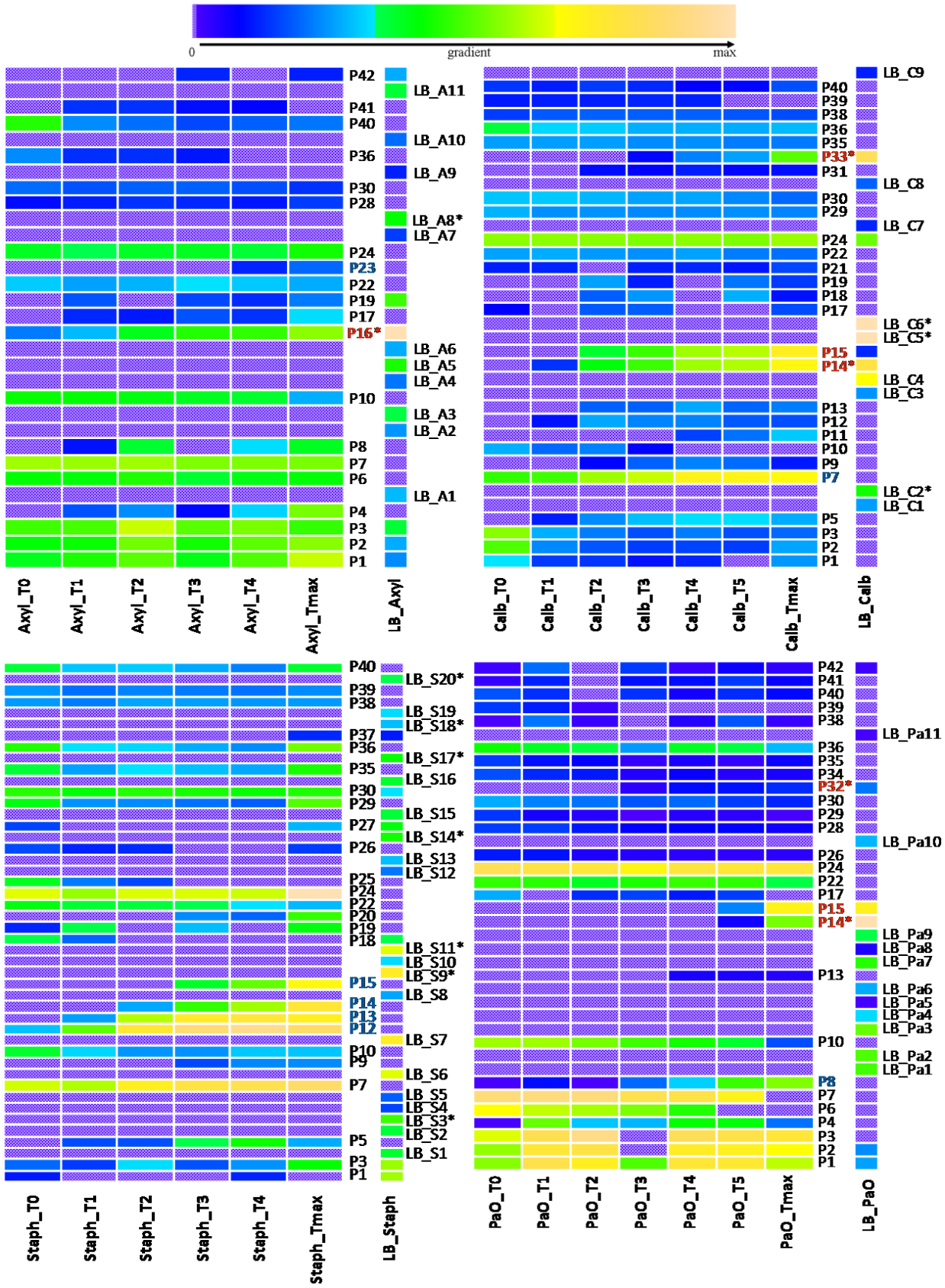


Figure 3: Heatmaps showing peak-integrals of chromatograms from *A. xylosoxidans* (**Axyl**), *C. albicans* (**Calb**), *S. aureus* (**Staph**) and *P. aeruginosa* (**PaO**). T0 - Tmax are time points of VOCs extraction from the host-pathogen interaction cultures. Peaks were aligned by their retention times as well as mass spectra and numbered

chronologically P1-42. In each heatmap, the last column shows peak compositions from corresponding monocultures on LB agar. Peaks are individually labelled and integrated in chronological order according to peaks P1-42 for each pathogen. Characteristic VOCs released in both cultures are marked in red, those released only in the interaction cultures are marked in blue. Additionally, peaks exhibiting similar retention times and mass spectra, in comparison with the monoculture biofilms, were highlighted with *-symbol.

To further verify the individual peak composition of each pathogen, VOCs extractions were performed during microbial growth on the human epithelial cell line A549. Growth of pathogens and status of the cell monolayer was observed visually by microscopy coupled with digital imaging (**Figure 2A, B**). A549 cells in interaction with *P. aeruginosa* showed increasing signs of damage, by means of decreasing confluency. No such damage was observed for interaction with the other pathogens. *C. albicans* was growing in hyphal form and appeared to be attached to the A549 cells, whereas *S. aureus* grew in its typical clusters. Peak profiles, taken from host-pathogen interaction cultures during the course of growth/interaction, are shown in **Figure 3**. Peak intensities were calculated individually for each heatmap, therefore, the representing colour gradients are not comparable between the different figures. All interaction cultures show a rather complex profile. Strong fluctuations of intensities are observed for various peaks throughout a series of VOCs extractions revealing no obvious pattern, whereas several other peaks did not show any major change in their intensities. Peaks P22, P24, P30 were present in all chromatograms with high intensities. Others were not detected in all cultures but gave constantly signals in a series of VOCs extractions for one or more interaction cultures (e.g. P38, P39, P30, P28, P29). However, major differences in the profile composition were observed for all host-pathogen interactions. In order to point out these differences and to weight their significance, samples where compared by their peak profiles using the Euclidean distance algorithm after transformation (log X+1), results are ordinated in the MDS plot in **Figure 4**. Logarithmic transformation was used to reduce the impact of large values from single peaks in order to keep the complete profile in consideration including peaks with relatively small values. Each time point was indicated by a symbol according to the pathogen, labelled T0 to Tmax and combined with a trajectory in chronological order. The MDS plot revealed an obvious change in the peak composition during incubation for each interaction culture. For all four pathogens, starting points (T0) were ordinated in the centre of the plot. During incubation, the statistical distances between the peak compositions of each sample was increasing and the time points are shifting away from each other to different quarters of the MDS plot, indicating a continuous change over time for each interaction culture.

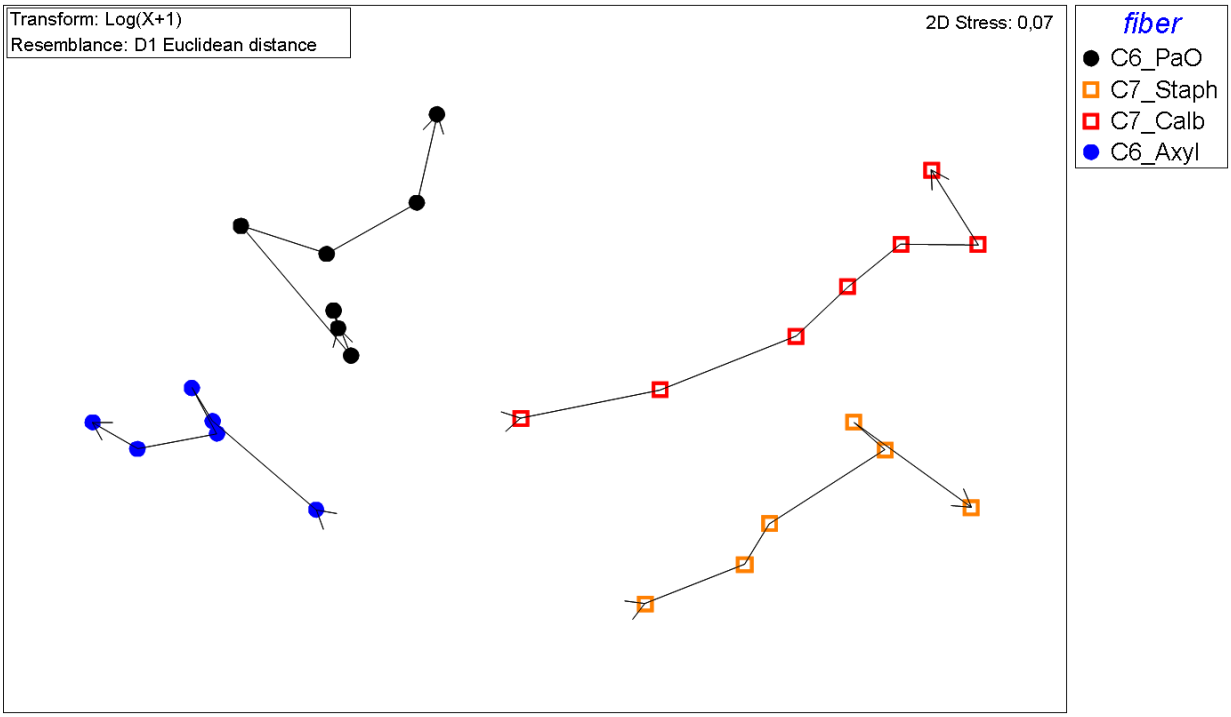


Figure 4: MDS plot of host-pathogen interaction cultures. Each symbol represents one time point of VOCs extraction. Distances between symbols reflect differences in the VOC compositions monitored for each sample and time point: the closer the symbols are to each other, the more similar were the VOC compositions. Symbols for *C. albicans* are in red, *S. aureus* in orange, *A. xylosoxidans* in blue and *P. aeruginosa* in black. Trajectories combine the different time points of each interaction culture.

B. Characteristic VOCs

Except for *C. albicans*, the trajectories revealed some discrepancies between the changing peak composition and the time line. For *P. aeruginosa* and *A. xylosoxidans* time points T1 and T3 were closer than T2 and T3. For *S. aureus* time points T3 and Tmax were closer than T4 and Tmax. In order to reveal which peaks were responsible for the main differences over time, it was hypothesized that peaks once they came up and show a steady increase should be considered to come from volatile compounds released in interaction between host and pathogen or by the growing pathogen alone. To all these peaks was further referred to as characteristic VOCs. Peaks P12 and P13 of *S. aureus* were also considered, although not continuously increasing but staying at high levels at later time points. In **Figure 3**, characteristic VOCs are marked according to comparisons with the corresponding monoculture on LB agar, shown in the last column of each heatmap. Peaks fulfilling the above criteria for characteristic VOCs and released in both, the interaction culture and the corresponding monoculture, are marked in red and those just released in the interaction culture are marked in blue. Conspicuous peaks already observed in the monoculture biofilms (**Figure 1A, B**) and also meeting peak integration criteria in this experiment were listed in

Table 2 and additionally highlighted in **Figure 3**. Signals constantly considered in all cultures and biofilms with the particular pathogen were P33 for *C. albicans*, P32 for *P. aeruginosa* and P14 for both of these microorganisms. Although ubiquitously present in all monoculture biofilms on LB agar, P16 was just observed in the host-pathogen interaction culture of *A. xylosoxidans*. Notably, none of the characteristic VOCs detected in the interaction culture with *S. aureus* was observed in the monoculture biofilm on LB agar.

Microorganism	Signals in biofilms	Signals in interaction cultures & corresponding monocultures
A. xylosoxidans	10.921 min	P16
	20.056 min	LB_A8
	22.241 min	/
C. albicans	3.196 min	LB_C2
	10.221 min	P14
	10.921 min	LB_C5/LB_C6
	22.581 min	P33
S. aureus	3.196 min	LB_S3
	10.921 min	LB_S9
	14.192 min	LB_S11
	19.811 min	LB_S14
	23.250 min	LB_S17
	25.536 min	LB_S18
	27.471 min	/
	27.553 min	LB_S20
	29.705 min	/
P. aeruginosa	10.211 min	P14
	10.921 min	/
	22.241 min	/
	22.456 min	P32
	24.833 min	/

Table 2: Matching conspicuous peaks detected for pathogens under different *in vitro* conditions. Signals from monoculture biofilms were given by their retention times in minutes (min) and signals from host-pathogen interaction cultures or corresponding monocultures were mentioned according to their labels in Figure 3. Peaks in one row were matching in the different cultures and biofilms whereas the slash indicates no matching peak within parameters of integration.

To further test the hypothesis of characteristic VOCs, a new MDS plot was calculated taken only these peaks into account (**Figure 5A**). The trajectories revealed that composition of characteristic VOCs was changing continuously in the samples throughout a series of VOC extractions. Each time point was more distant to T0 than the previous ones. One exception has

to be mentioned here: the first three time points (T0-T2) in *P. aeruginosa* were overlapping and therefore showed no distance to each other at all. For all microorganisms characteristic VOCs did make the main differences during incubation. The final composition of characteristic VOCs from each host-pathogen interaction culture, one day post infection, were shown in **Figure 5B**. Several of these VOCs were detected for more than one pathogen but their intensities did vary when compared with other peaks released in the same interaction culture. Signals P13, P14, P15 were detected in all cultures, except for *A. xylosoxidans* interaction cultures, and P7 was likewise only not detected for *P. aeruginosa*. Characteristic VOCs unique to one interaction culture are P16 and P23 for *A. xylosoxidans*, P11 and P33 for *C. albicans* and P32 for *P. aeruginosa*. *S. aureus* did not show any unique signal in the host-pathogen interaction cultures but different relative intensities of the emitted VOCs in comparison to other the host-pathogen cultures.

Summing up, not just presence and absence of released VOCs but also the relative intensities were characteristic for each host-pathogen interaction culture, allowing discrimination between the different pathogens *in vitro* and even between different time points during incubation of the interaction cultures.

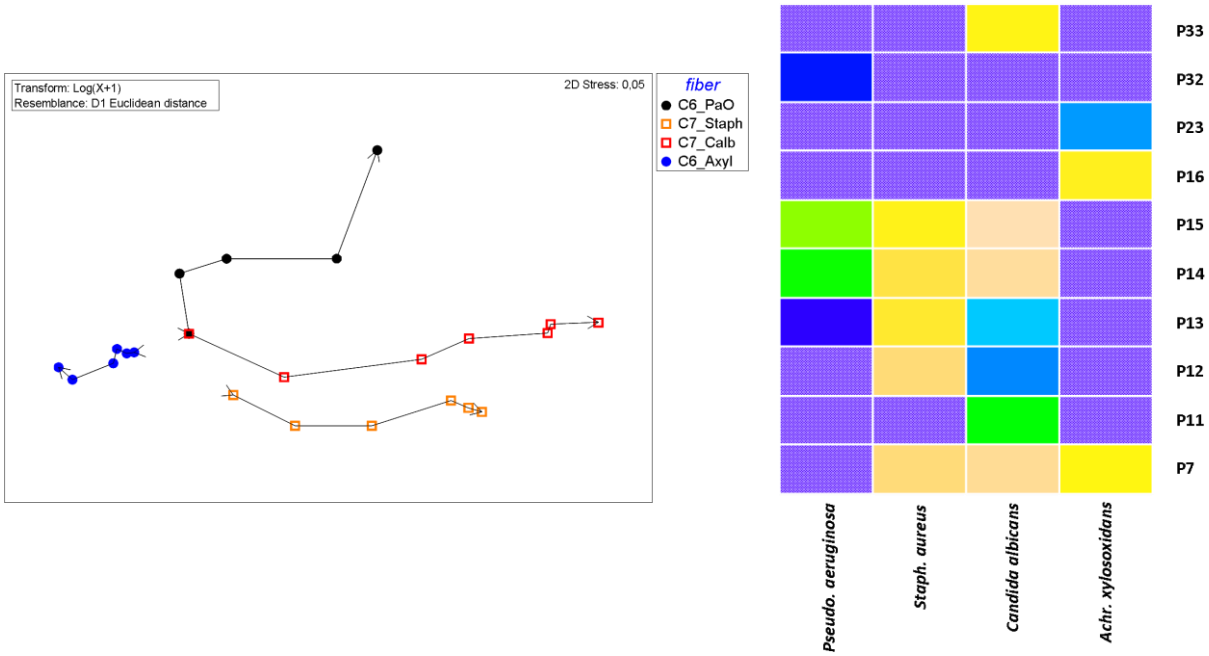


Figure 5A, B: Discriminative characteristic VOC compositions**. A.** MDS plot was calculated just considering discriminative characteristic VOCs of each sample. Trajectories combine the series of VOCs extractions from each host-pathogen interaction culture. **B.** The heatmap was calculated with characteristic VOCs detected in the last time points Tmax. Relative intensities of peaks are indicated by the colour gradient.

4.4.3 *In vivo* VOCs monitoring in exhaled breath of CF patients

Since volatile organic compounds allowed discrimination between major CF pathogens *in vitro*, the next step was to investigate the diagnostic potential of these characteristic VOCs for clinical diagnostic purposes *in vivo*. Therefore, a breath analytical procedure was established based on the same principles like the *in vitro* VOCs extractions. After adaptation to clinical needs, this method was integrated in the routine ambulance examinations for CF patients. Furthermore, the individual microbiome was elucidated by culture independent methods for each patient, taken at same day as the exhaled air samples, thus allowing correlation between both clinical samples. Comparable peak profiles were generated for nine CF patients and two healthy individuals. Additionally, four air room samples were taken and three profiles were generated from SPME fibers stored in the sampling locations without exposition. Identification under defined parameter revealed in total 36 peaks for all profiles showing differences in retention times and mass spectra. They were numbered hP1-hP36 and the last peak was assessed at a retention time of 26.739 min (**Figure 6**). Again, rather complex profiles for all extractions were observed. Remarkable peaks were hP3 and hP5, which were present in high intensities in all exhaled breath extractions, whereas controls showed no signals or just in low intensities. Peaks hP4 and hP6 were detected in all samples extracted at the clinic, but never in those made at the institute. Signal hP10 was ubiquitously present in all profiles and intensities were higher in all samples stored or transported for more than 1 hour after exposition. Furthermore, SPME fibers showed individual peak patterns detected in all chromatographic runs made with one specific fiber (e.g. hP32 and hP25 for fiber #7). Several peaks were just detected in one or more profiles of exhaled breath from CF patients, although mostly with low intensities. Since none of these peaks were matching with the characteristic VOCs or conspicuous individual VOCs of the *in vitro* extractions, by means of retention times and mass spectra, the exhaled air profiles were further analysed beyond the previously defined parameters of peak integration. Using the MS Data Review, chromatographic runs from *in vivo* VOCs extractions were re-examined by displaying chromatograms generated by the particular mass-charge ratios (*m/z*) of the discriminative VOCs detected in the *in vitro* experiments. Thus, clean chromatograms for those compounds could have been provided out of the full scan runs giving a higher resolution for the peaks of interest. However, none of the discriminative VOCs were detected in the exhaled breath profiles.

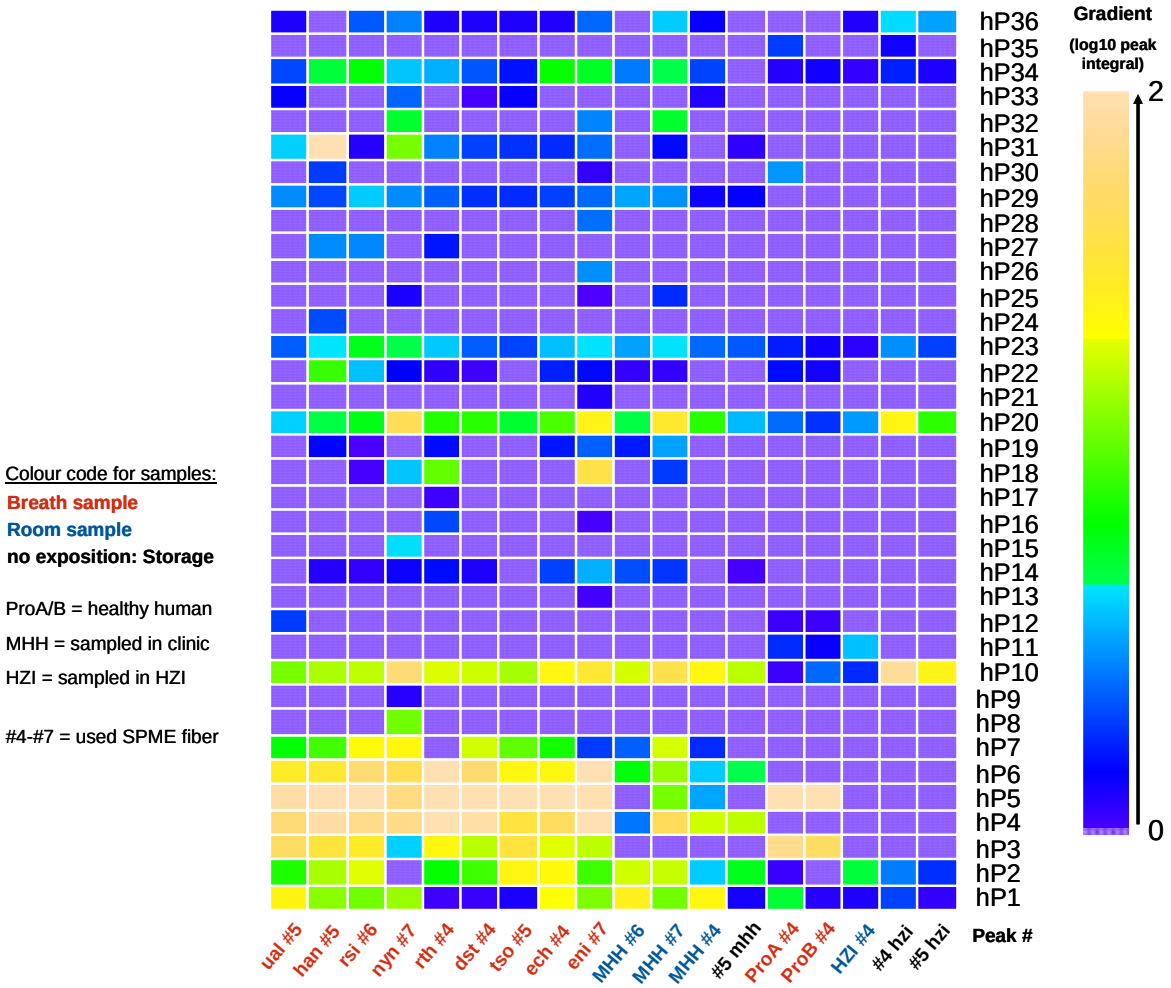


Figure 6: VOCs composition from exhaled air of CF patients. VOCs extractions from exhaled breath are marked in red, whereupon ProA / ProB are healthy non-CF individuals. Room samples are indicated in blue and SPME fibers without exposition are marked in black. Peaks are identified under given parameter and numbered hP1-hP36. The colour gradient is indicating the peak intensities relative to each other.

For a detailed analysis of the peak profiles, resemblance matrices were calculated with *in vivo* data and samples were accordingly ordinated in a MDS plot (**Figure 7**). Although none of the characteristic VOCs or conspicuous individual peaks was detected, the MDS plot revealed a clear clustering of all exhaled breath profiles, clearly separating the healthy individuals from the CF patients. Similarly, all control samples ordinated closely together. Furthermore, the MDS plot showed a separation between samples extracted in the clinic (grey box) and those extracted at the institute. Comparisons with clinical or microbial data sets from the patients revealed no further correlations explaining the pattern in the MDS plot. Based on the assumption that distances in the resemblance matrix were dominated by a few peaks that were independent from microbial or clinical data sets, "BEST" routine analysis was performed on the distance matrix in order to identify the main peaks, responsible for the ordination pattern shown in the MDS plot, with a maximum number of five variables. Thus every peak, given in

a subset of five with a Spearman rank coefficient higher than 0.940, was considered. A new distance matrix was calculated without the dominant peaks hP3-hP6, hP8, hP20, hP31 and displayed in a new MDS plot (data not shown). Again, comparisons between microbial and clinical data with the pattern in the MDS plot revealed no correlation.

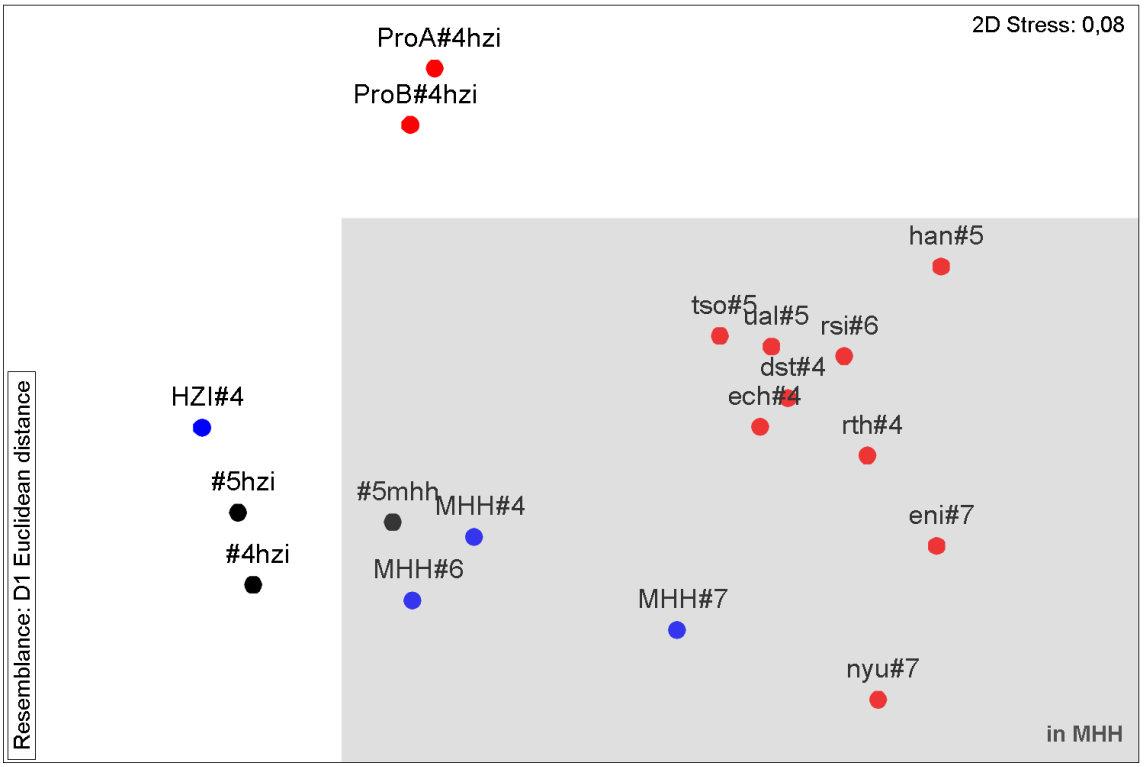


Figure 7: MDS plot of VOCs extraction from *in vivo* experiments. Peak compositions of different samples were compared: human exhaled air is marked in red, whereupon ProA / Pro9B are healthy non-CF individuals. Room samples are indicated in blue and SPME-fibers without exposition are marked in black. The numbers indicate each of the four individual SPME fibers of the type 75 µm Carboxen/PDMS, which were used for the extractions. Locations of sampling were mentioned with MHH/mhh for clinical samples and HZI/hzi for extractions performed in the institute.

4.5 Discussion

In this study, the emitted VOCs of different microorganisms from the vast spectrum of microbial pulmonary pathogens were investigated to assess if they could be used for species discrimination and further clinical diagnostic purposes. *In vitro* experiments with different growth conditions were combined with *in vivo* exhaled breath analysis of CF patients suffering from microbial respiratory tract colonization. Microorganisms included in this study were the major CF pathogens *P. aeruginosa* and *S. aureus*, the emerging innately multi-resistant betaproteobacterium *A. xylosoxidans* and the polymorphic fungus *C. albicans*.

4.5.1 Relevance of appropriate statistical methods in VOC emission analysis

Fast developing bio-analytical technologies such as the extraction and analysis of microbial VOCs *in vitro* and exhaled VOCs *in vivo* are generating large data sets requiring advanced mathematical tools for comparative studies. Thus, bio-statistical methods are of increasing importance in VOCs profiling with a particular emphasis on data exploration by visualization and identification of grouping trends (Boots et al. 2012). The multidimensional scaling (MDS) plots used in this study have the advantages to feature multivariant similarity measurements and simultaneously allow information about the variables that are responsible for certain groupings. In this study, the rather complex VOC emission profiles generated *in vitro* and *in vivo* were simplified by reducing the information to Euclidean distances between all samples. The resulting MDS plots clearly revealed discriminative VOC emissions, enable differentiation between species *in vitro* and even visualized a specific diversification as a function of time. Similarly, grouping of *in vivo* exhaled breath samples clearly separated from control samples could be illustrated. Multivariant analysis using (distance) resemblance matrices and MDS plots are therefore an appropriate statistical tool to analyze and compare VOC profiles. This is in accordance with another recent study using this statistical strategy to differentiate between microbial VOCs emissions *in vitro* (Thorn et al. 2011).

4.5.2 Application of *in vitro* VOC emission models for microbial discrimination

In vitro microbial VOCs profiles were studied and compared using two different model systems: (i) monoculture biofilms and (ii) host-pathogen interaction cultures. Biofilm formation is considered to be a major factor in recurrent and chronic pulmonary infections, especially in CF patients (Liu & Post 2009). Rather simple monoculture biofilms were grown *in vitro* and used to verify the potential of VOCs to discriminate between different pathogens. To achieve conditions more similar to the human respiratory tract, the host-pathogen

interaction culture system was established. To this end, the human lung epithelial cell monolayer was inoculated with high amounts of pathogens to mimic infection conditions. In bronchoalveolar lavage (BAL) studies a threshold to discriminate infection from colonization was defined as 10^4 or 10^5 CFU ml^{-1} (American Thoracic Society & Infectious Diseases Society of America 2005), which is well below the numbers used in this study even when taken the dilution factor of BAL diagnostics into account. Only the A549 monolayer inoculated with *P. aeruginosa* showed visible signs of damage but all pathogens showed typical growing behaviours, therefore this system was considered to be more an interaction rather than infection culture. VOC extractions from biofilms showed individual emissions, allowing differentiation between the pathogenic species included in this study. The host-pathogen interaction cultures revealed a similar picture: VOC emissions allow differentiation of species and even of different time points during incubation. However, most of the detected volatiles were released by more than one microorganism. Studies in the past concentrated on the search of single biomarker molecules for a certain respiratory pathogen (Chambers et al. 2009) (Phillips et al. 2010) (Scott-Thomas et al. 2010a), which is accompanied by several difficulties that are further discussed below in the context of this study. Comparisons of profiles from the same pathogen in biofilm and host-pathogen interaction culture showed major differences. Only for *P. aeruginosa* and *C. albicans*, one unique VOC could be observed in all experimental set-ups. The profiles extracted in biofilms and corresponding monocultures showed little differences as well, these compounds had been detected but were not fulfilling the parameter of peak integration and were therefore not considered in the profiles of the corresponding monocultures. *S. aureus* is of particular interest, since rather different volatile metabolites were observed in the biofilm and during various time points of interaction with A549 cells. This might be due to different physiological states of the bacterium during VOCs extractions and profiles might get more similar the longer they are incubated (O'Hara & Mayhew 2009) or due to the different atmospheres with unequal oxygen and CO_2 levels. Since cultures on LB agar grew in closed vials, oxygen supply is limited and this might influence the metabolism and production of VOCs in particular. Transcriptional and metabolic changes in *S. aureus*, adapting to different oxygen concentrations were well described in several studies (Fuchs et al. 2007) (Sun et al. 2012).

For all microorganisms in this study, certain differences in the VOC emissions were observed when biofilm and interaction cultures were compared. Furthermore, VOCs emissions were shown as a function of time during interaction with host cells with variable but specific compositions for each microorganism. It was previously shown that production of VOCs is

directly linked with the physiological state of bacteria or fungi and therefore, temporal variability of VOC emissions have to be taken into account for the search of volatile marker compounds (Bunge et al. 2008). Furthermore, the substrates available for the microorganisms also alter the VOC production, intensity and even the type of compound are strongly influenced by the growth media (Scotter et al. 2005). Several studies investigated VOCs released by *P. aeruginosa* and *S. aureus in vitro* (Cox & Parker 1979) (Labows et al. 1980) (Zechman & Labows Jr. 1985) (Barbieri et al. 2005) (Carroll et al. 2005). Different culture conditions, extraction methods and analysis strategies revealed numerous compounds in the headspace of these microorganisms with partly divergent results (Allardyce et al. 2006a) (Preti et al. 2009) (Zhu et al. 2010) (Shestivska et al. 2011) (Goeminne et al. 2012) (Filipiak et al. 2012). Furthermore, since metabolic pathways underlie similar principles throughout all microorganisms other sources for several microbial VOCs are found including mammalian breath (Korpi et al. 2009). Recently, 2-aminoacetophenone was getting in the focus as biomarker molecule for *P. aeruginosa* and was used in a first breath analysis study to discriminate *P. aeruginosa*-colonized CF patients from others (Scott-Thomas et al. 2010b). However, other studies detected this compound also in the headspace of different respiratory pathogens such as *S. aureus* and *Burkholderia cepacia* (Allardyce et al. 2006b) (Thorn et al. 2011). Taken together, results of this study go in accordance with the literature showing that identification of reliable single molecule biomarker *in vitro*, unique for each pathogen, seemed to be highly disputable. Growth of microorganisms was observed *in vitro* by their specific VOC compositions. Only for *P. aeruginosa* and *C. albicans* unique compounds released under all tested conditions were observed, but their species-specificity and universal emission under the specific conditions in the human lung remains to be demonstrated.

4.5.3 Feasibility of exhaled VOCs analysis for diagnostics of pulmonary infections

None of the VOCs from *in vitro* cultures were detected in exhaled breath of CF patients although colonization of patients with pathogens was assessed. This highlights the importance of performing studies in humans to identify true biomarkers. Difficulties in the search for single volatile biomarker are discussed in detail above. Further explanations, could be very low concentrations of the compounds and the relatively healthy status of the studied individuals from the CF ambulance centre. For SPME, an individual detection limit for each compound has to be considered. Exhaled human endogenous VOCs are considered to be present in levels of parts-per-billion (ppb) with higher levels of metabolites associated with diseased states (Amann et al. 2007). Also a possible overrepresentation of VOCs from the

ambient air has to be considered. Further data mining procedures like the search for particular mass-charge ratios of microbial VOCs or "BEST" routine analysis to exclude ambient air or ubiquitous human endogenous VOCs did not reveal any positive correlations. Far more than 500 different VOCs from diverse origins and biochemical pathways have been detected with various methods in human exhaled air (Miekisch et al. 2004). Just small subsets of them were detected in this study, revealing already complex profiles for each sample. Data exploration by MDS plots revealed well defined groupings of breath samples and controls. However, a strong influence by the ambient air was observed and may have a strong influence on the separation of exhaled air samples of healthy individuals and CF patients. In a comparative study investigating concentrations of 12 VOCs in exhaled breath from 20 CF patients versus 20 healthy individuals, slightly altered compositions were observed between both groups and gave no unambiguous correlations (Barker et al. 2006). Barker et al. also observed increased *n*-pentane levels, which are discussed as a marker for inflammatory diseases, including lung diseases (Olopade et al. 1997) (Schubert et al. 1998). However, a strong correlation between room and exhaled *n*-pentane was likewise observed, suggesting to take increased hydrocarbon levels cautiously (Gorham et al. 2009). Nevertheless, in this study differences in exhaled air profiles were observed and resulted in a certain grouping of VOCs profiles from CF patients. To assess further correlations with the entire individual microbiome, medications or health conditions, the cohort needs to be expanded. Robroeks et al. recently considered 1099 VOCs in the breath of paediatric CF patients and achieved 100% identification of CF patients and healthy individuals by using 22 VOCs (Robroeks et al. 2010). Even further discrimination of *P. aeruginosa* colonized patients was achieved by the use of 14 compounds. These results strongly support the hypothesis that rather a composition of exhaled compounds than single molecules might allow the differentiation between respiratory infections. To validate the suitability and importance of VOCs for the purpose of exhaled breath diagnostics, their biosynthetic pathways and origins have to be further elucidated (Miekisch et al. 2004) (Buszewski et al. 2008). For the methodology, a key issue is the breath sampling and exhaled VOCs extraction procedure. On the one hand it has to be easy-to-use for clinical purposes and on the other hand cross-contaminations have to be avoided. A similar strategy with the commercial BIO-VOC sampler and SPME fiber was also used in other diagnostic research approaches for respiratory diseases (Poli et al. 2005) (Poli et al. 2010). To bind exhaled VOCs, the choice of the adsorbent material on the sorbent trap strongly depends on the compounds of interest. Caldeira et al. further compared different sampling strategies and showed advantages of Tedlar® bags in terms of number and peak area of collected compounds

(Caldeira et al. 2011). However, their handling and cleaning requires more training making an implementation in clinics more difficult.

4.6 Conclusion and perspectives

Exhaled VOCs analysis for the detection of pulmonary infectious diseases is still in the research phase but evidence is accumulating that it could be a rapid and efficient diagnostic tool in the future. This study showed that the presence and relative intensities of VOCs allowed differentiation of microbial species and support the hypothesis that the composition of exhaled VOCs in human breath delivers information to identify patients with certain infections. A major challenge will be the development of a standard breath sampling and VOCs extraction method suitable for clinical diagnostics. Statistical methods, like MDS plots, are essential tools for this purpose to measure multivariant similarities and to visualize correlation patterns. The more exhaled VOCs samples are considered the more powerful these statistic approaches will be.

4.7 References

Allardyce RA, Hill AL & Murdoch DR (2006a) The rapid evaluation of bacterial growth and antibiotic susceptibility in blood cultures by selected ion flow tube mass spectrometry. *Diagn. Microbiol. Infect. Dis.* 55: 255–261.

Allardyce RA, Langford VS, Hill AL & Murdoch DR (2006b) Detection of volatile metabolites produced by bacterial growth in blood culture media by selected ion flow tube mass spectrometry (SIFT-MS). *Journal of Microbiological Methods* 65: 361–365.

Amann A, Spanel P & Smith D (2007) Breath Analysis: The Approach Towards Clinical Applications. *Mini Reviews in Medicinal Chemistry* 7: 115–129.

American Thoracic Society & Infectious Diseases Society of America (2005) Guidelines for the Management of Adults with Hospital-acquired, Ventilator-associated, and Healthcare-associated Pneumonia. *American Journal of Respiratory and Critical Care Medicine* 171: 388–416.

Barbieri E, Gioacchini AM, Zambonelli A, Bertini L & Stocchi V (2005) Determination of microbial volatile organic compounds from Staphylococcus pasteuri against Tuber borchii using solid-phase microextraction and gas chromatography/ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry* 19: 3411–3415.

Barker M, Hengst M, Schmid J, Buers H-J, Mittermaier B, Klemp D & Koppmann R (2006) Volatile organic compounds in the exhaled breath of young patients with cystic fibrosis. *Eur. Respir. J.* 27: 929–936.

Boots AW, van Berkel JJBN, Dallinga JW, Smolinska A, Wouters EF & van Schooten FJ (2012) The versatile use of exhaled volatile organic compounds in human health and disease. *J Breath Res* 6: 027108.

Brakhage AA & Schroeckh V (2011) Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genet. Biol.* 48: 15–22.

Bunge M et al. (2008) On-Line Monitoring of Microbial Volatile Metabolites by Proton Transfer Reaction-Mass Spectrometry. *Appl Environ Microbiol* 74: 2179–2186.

Buszewski B, Ulanowska A, Ligor T, Jackowski M, Klodzinska E & Szeliga J (2008) Identification of volatile organic compounds secreted from cancer tissues and bacterial cultures. *Journal of Chromatography B* 868: 88–94.

Caldeira M, Barros AS, Bilelo MJ, Parada A, Câmara JS & Rocha SM (2011) Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J Chromatogr A* 1218: 3771–3780.

Carroll W, Lenney W, Wang T, Španěl P, Alcock A & Smith D (2005) Detection of volatile compounds emitted by Pseudomonas aeruginosa using selected ion flow tube mass spectrometry. *Pediatric Pulmonology* 39: 452–456.

Chambers ST, Bhandari S, Scott-Thomas A & Syhre M (2011) Novel diagnostics: progress toward a breath test for invasive Aspergillus fumigatus. *Med. Mycol.* 49 Suppl 1: S54–61.

Chambers ST, Syhre M, Murdoch DR, McCartin F & Epton MJ (2009) Detection of 2-pentylfuran in the breath of patients with Aspergillus fumigatus. *Med. Mycol.* 47: 468–476.

Cox CD & Parker J (1979) Use of 2-aminoacetophenone production in identification of Pseudomonas aeruginosa. *J. Clin. Microbiol.* 9: 479–484.

Dunkel M et al. (2009) SuperScent--a database of flavors and scents. *Nucleic Acids Research* 37: D291–D294.

Dyne D, Cocker J & Wilson HK (1997) A novel device for capturing breath samples for solvent analysis. *Sci. Total Environ.* 199: 83–89.

Filipiak W, Sponring A, Baur MM, Filipiak A, Ager C, Wiesenhofer H, Nagl M, Troppmair J & Amann A (2012) Molecular analysis of volatile metabolites released specifically by staphylococcus aureus and pseudomonas aeruginosa. *BMC Microbiol.* 12: 113.

Fuchs S, Pané-Farré J, Kohler C, Hecker M & Engelmann S (2007) Anaerobic Gene Expression in Staphylococcus aureus. *J Bacteriol* 189: 4275–4289.

Goeminne PC, Vandendriessche T, Van Eldere J, Nicolai BM, Hertog ML & Dupont LJ (2012) Detection of Pseudomonas aeruginosa in sputum headspace through volatile organic compound analysis. *Respiratory Research* 13: 87.

Gorham KA, Sulbaek Andersen MP, Meinardi S, Delfino RJ, Staimer N, Tjoa T, Rowland FS & Blake DR (2009) Ethane and n-pentane in exhaled breath are biomarkers of exposure not effect. *Biomarkers* 14: 17–25.

Harrison F (2007) Microbial ecology of the cystic fibrosis lung. *Microbiology (Reading, Engl.)* 153: 917–923.

Henderson KA & Matthews IP (2002) Biological monitoring of midwives' exposure to N2O using the Bio-VOC breath sampler. *Journal of Exposure Science and Environmental Epidemiology* 12: 309–312.

Kai M, Haustein M, Molina F, Petri A, Scholz B & Piechulla B (2009) Bacterial volatiles and their action potential. *Applied Microbiology and Biotechnology* 81: 1001–1012.

Korpi A, Järnberg J & Pasanen A-L (2009) Microbial volatile organic compounds. *Crit. Rev. Toxicol.* 39: 139–193.

Kramer R & Abraham W-R (2012) Volatile sesquiterpenes from fungi: what are they good for? *Phytochemistry Reviews* 11: 15–37.

Labows JN, McGinley KJ, Webster GF & Leyden JJ (1980) Headspace analysis of volatile metabolites of Pseudomonas aeruginosa and related species by gas chromatography-mass spectrometry. *J. Clin. Microbiol.* 12: 521–526.

Liddell K (1976) Smell as a diagnostic marker. *Postgrad Med J* 52: 136–138.

Liu Y-CC & Post JC (2009) Biofilms in pediatric respiratory and related infections. *Curr Allergy Asthma Rep* 9: 449–455.

Miekisch W, Schubert JK & Noeldge-Schomburg GF. (2004) Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clinica Chimica Acta* 347: 25–39.

O'Brien J & Wright GD (2011) An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* 22: 552–558.

O'Hara M & Mayhew CA (2009) A preliminary comparison of volatile organic compounds in the headspace of cultures of Staphylococcus aureus grown in nutrient, dextrose and brain heart bovine broths measured using a proton transfer reaction mass spectrometer. *J Breath Res* 3: 027001.

Olopade CO, Zakkar M, Swedler WI & Rubinstein I (1997) Exhaled pentane levels in acute asthma. *Chest* 111: 862–865.

Pawliszyn J, ed. (2009) *Handbook of Solid Phase Microextraction*. Chemical Industry Press of China http://www.spme.uwaterloo.ca/spmedb/SPME_book_leaflet.pdf (Accessed November 7, 2012).

Phillips M, Basa-Dalay V, Bothamley G, Cataneo RN, Lam PK, Natividad MPR, Schmitt P & Wai J (2010) Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb)* 90: 145–151.

Phillips M, Cataneo RN, Condos R, Ring Erickson GA, Greenberg J, La Bombardi V, Munawar MI & Tietje O (2007) Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis (Edinb)* 87: 44–52.

Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, Bianchi L, Rusca M & Mutti A (2005) Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respir. Res.* 6: 71.

Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, Casalini A & Mutti A (2010) Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 878: 2643–2651.

Preti G, Thaler E, Hanson CW, Troy M, Eades J & Gelperin A (2009) Volatile compounds characteristic of sinus-related bacteria and infected sinus mucus: analysis by solid-phase microextraction and gas chromatography-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 877: 2011–2018.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. http://www.R-project.org (Accessed December 13, 2012).

Robroeks CMHHT et al. (2010) Metabolomics of Volatile Organic Compounds in Cystic Fibrosis Patients and Controls. *Pediatric Research* 68: 75–80.

Schubert JK, Müller WP, Benzing A & Geiger K (1998) Application of a new method for analysis of exhaled gas in critically ill patients. *Intensive Care Med* 24: 415–421.

Schulz S & Dickschat JS (2007) Bacterial volatiles: the smell of small organisms. *Nat Prod Rep* 24: 814–842.

Scotter JM, Langford VS, Wilson PF, McEwan MJ & Chambers ST (2005) Real-time detection of common microbial volatile organic compounds from medically important fungi by Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS). *J. Microbiol. Methods* 63: 127–134.

Scott-Thomas A, Syhre M, Pattemore P, Epton M, Laing R, Pearson J & Chambers S (2010a) 2-Aminoacetophenone as a potential breath biomarker for Pseudomonas aeruginosa in the cystic fibrosis lung. *BMC Pulmonary Medicine* 10: 56.

Scott-Thomas A, Syhre M, Pattemore P, Epton M, Laing R, Pearson J & Chambers S (2010b) 2-Aminoacetophenone as a potential breath biomarker for Pseudomonas aeruginosa in the cystic fibrosis lung. *BMC Pulmonary Medicine* 10: 56.

Shestivska V, Nemec A, Dřevínek P, Sovová K, Dryahina K & Španěl P (2011) Quantification of methyl thiocyanate in the headspace of Pseudomonas aeruginosa cultures and in the breath of cystic fibrosis patients by selected ion flow tube mass spectrometry. *Rapid Communications in Mass Spectrometry* 25: 2459–2467.

Shirasu M & Touhara K (2011) The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J Biochem* 150: 257–266.

Sun J-L, Zhang S-K, Chen J-Y & Han B-Z (2012) Metabolic profiling of Staphylococcus aureus cultivated under aerobic and anaerobic conditions with (1)H NMR-based nontargeted analysis. *Can. J. Microbiol.* 58: 709–718.

Syhre M & Chambers ST (2008) The scent of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* 88: 317–323.

Syhre M, Manning L, Phuanukoonnon S, Harino P & Chambers ST (2009) The scent of Mycobacterium tuberculosis--part II breath. *Tuberculosis (Edinb)* 89: 263–266.

Syhre M, Scotter JM & Chambers ST (2008) Investigation into the production of 2-Pentylfuran by Aspergillus fumigatus and other respiratory pathogens in vitro and human breath samples. *Med. Mycol.* 46: 209–215.

Teranishi R, Mon TR, Robinson AB, Cary P & Pauling L (1972) Gas chromatography of volatiles from breath and urine. *Anal. Chem.* 44: 18–20.

Thomas AN, Riazanskaia S, Cheung W, Xu Y, Goodacre R, Thomas CLP, Baguneid MS & Bayat A (2010) Novel noninvasive identification of biomarkers by analytical profiling of chronic wounds using volatile organic compounds. *Wound Repair and Regeneration* 18: 391–400.

Thorn RMS, Reynolds DM & Greenman J (2011) Multivariate analysis of bacterial volatile compound profiles for discrimination between selected species and strains in vitro. *Journal of Microbiological Methods* 84: 258–264.

UWE Bristol (2012) Learning to sniff bad breath can help diagnose oral disease. http://info.uwe.ac.uk/news/uwenews/news.aspx?id=2335 (Accessed November 22, 2012).

Woolfenden E (2010a) Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 2. Sorbent selection and other aspects of optimizing air monitoring methods. *Journal of Chromatography A* 1217: 2685–2694.

Woolfenden E (2010b) Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air: Part 1: Sorbent-based air monitoring options. *Journal of Chromatography A* 1217: 2674–2684.

Zechman JM & Labows Jr. JN (1985) Volatiles of Pseudomonas aeruginosa and related species by automated headspace concentration – gas chromatography. *Canadian Journal of Microbiology* 31: 232–237.

Zhu J, Bean HD, Kuo Y-M & Hill JE (2010) Fast Detection of Volatile Organic Compounds from Bacterial Cultures by Secondary Electrospray Ionization-Mass Spectrometry. *Journal of Clinical Microbiology* 48: 4426–4431.

Zlatkis A, Brazell RS & Poole CF (1981) The role of organic volatile profiles in clinical diagnosis. *Clin. Chem.* 27: 789–797.

Zlatkis A, Lichtenstein HA & Tishbee A (1973) Concentration and analysis of trace volatile organics in gases and biological fluids with a new solid adsorbent. *Chromatographia* 6: 67–70.

4.8 Supplementary Material

patient	sampling date	gender	date of birth	FEV_1 value	clinical diagnostics
tso	01.11.2011	♀	03.06.1975	/	*P. aeruginosa; Candida albicans*
dst	02.11.2011	♂	29.08.1966	36 %	*P. aeruginosa; Candida albicans*
han	15.11.2011	♂	17.11.1979	56 %	*P. aeruginosa; S. aureus; Penicillium*
rth	15.11.2011	♂	31.12.1970	46 %	*P. aeruginosa; S. aureus; Aspergillus fumigatus*
rsi	16.11.2011	♀	05.12.1968	/	*P. aeruginosa; Yeast species*
ech	23.11.2011	♂	23.06.1963	25 %	*P. aeruginosa; Pseudallescheria boydii*
eni	23.11.2011	♀	08.04.1970	19 %	*P. aeruginosa; Streptococci*
nyu	06.12.2011	♂	19.06.1981	29 %	*P. aeruginosa; S. aureus; Candida albicans*
ual	07.12.2011	♂	20.07.1978	26 %	*P. aeruginosa; S aureus; Candida albicans; Enterococcus spec.*

Table S1: Clinical data of CF patients reported from the CF outpatient clinic of the MHH. Anonymized identification code for each patient is given. FEV_1 value (forced expiratory volume in 1 second; predicted) indicates lung function for the respective patient: FEV_1 70 - 89 % shows mild lung malfunction, 40 - 69 % indicates moderate lung malfunction and less than 40 % is a sign of severe lung malfunction. Microorganisms detected in the individual sputum samples by the routine clinical diagnostics are mentioned.

Chapter V
General Discussion

5.1 Diagnostic methods for lower respiratory infections

Diagnostics of infectious diseases are critical to initiate the appropriate treatment. Clinical microbiological laboratories have already undergone important changes in the past decades and further improvements are needed to cope with new challenges and aspects elucidated in respective research, such as microbial community profiling (Raoult et al. 2004). For lower respiratory infections in CF, detection and monitoring of pathogens over time are key issues in health management. Clinical microbial diagnostics for CF lung diseases are based on cultivation methods. Though, in comparison with next generation sequencing (NGS), high accuracy was demonstrated for the detection of major pathogens, it only allowed a limited view of the community in terms of richness and diversity (Chapter II). Interestingly, profound contamination of sputum samples by oral cavity bacteria were not observed and the same species were found in lung tissue and in sputum taken prior to the lung explantation (Rogers et al. 2006) (Rudkjøbing et al. 2011). Therefore, communities in sputum samples assumingly display well the lower respiratory communities. Microbial community profiling, however, has attracted more interest in the past years by using a variety of approaches (Bittar & Rolain 2010). Hereby, culture-independent methods are necessary to explore the diversity of the communities and to establish compositional analyses (Rogers et al. 2009). Although well established in research, conventional fingerprinting methods are technically too complicated to be applied in routine diagnostics. However, they facilitate accurate profiling of the abundant species in complex communities for research purposes (Chapter II) (Chapter III) (Pedrós-Alió 2012). Commercially available NGS platforms may facilitate the translation of microbial profiling into clinical diagnostics. NGS technologies, like Illumina, offer high reproducibility and are capable to analyse large numbers of samples (Chapter II). However, drawbacks of these methods are time-consuming sample preparation and the bioinformatic effort. These issues substantially complicate the translation into routine diagnostics. Additionally for all sputum-based methods, it applies that bias may occur due to uneven

spatial distribution of certain bacteria in the sample and in the different compartments of the human lungs (Willner et al. 2012) (Goddard et al. 2012).

Analytical, sputum-independent diagnostic approaches, like the detection of volatile microbial biomarker in exhaled breath, may represent a future alternative in diagnostics (Chapter IV). In particular, volatile organic compounds (VOCs) were extracted from exhaled air and their potential to differentiate subgroups of CF patients by microbial infections was elucidated (Robroeks et al. 2010). Monitoring and discrimination of different pathogens *in vitro* was successfully approached by extraction and analyses of emitted microbial VOCs (Chapter IV). However, this approach implies characterization of specific compounds which might be a major drawback in comparison to NGS. Molecular microbial profiling methods, including fingerprinting, reveal the entire structure and composition of microbial communities without previous selection (Chapter II) (Chapter III). Discrimination of specific pathogens requires knowledge about respective marker compounds beforehand. In this term, a selection has to be done in exhaled breath analysis. Nevertheless, detection of universal inflammatory markers, which might include VOCs, may be used for indirect detection of infectious agents (Paredi et al. 1999) (Paredi et al. 2000). Finally, the characterization of exhaled endogenous or microbial VOCs, including knowledge about their origins or even ecological functions, will contribute to a more profound and comprehensive understanding of lower respiratory tract infections and reinforce the great potential of exhaled air analyses for diagnostic purposes.

5.2 Microbial community structure and composition in CF sputum

Microbial communities observed in sputum samples from CF patients consist mostly only of a few species, whereas the entire community observed in a cohort appears to be more complex with greater variation between patients (Chapter II) (Chapter III). Therefore, each community that colonizes the lower respiratory tract might be considered to be rather unique (Stressmann et al. 2012). The principles for this specification are following a universal pattern: in a review, focussing on marine microbial communities, Fuhrmann (Fuhrman 2009) discussed the classic dictum about microbial distribution “everything is everywhere” and that the environment selects those organisms best fitted to the individual conditions in one site. Assumingly, the same principle applies in the lower respiratory tracts of CF patients. The observation, that the more widespread bacteria are also the most abundant ones in individual samples was made in environmental microbiology and recently also in CF sputum (Pommier et al. 2007) (Guss et al. 2011) (Zhao et al. 2012). This phenomenon was becoming more apparent for CF

respiratory tracts with structural and compositional analyses in a large number of patients (Chapter II) (Chapter III). Assumingly, abundant species are more likely to be detected whereas rare bacteria might mostly be present in samples but more likely under the detection limit (Fuhrman 2009). Only in certain cases, these rare bacteria find an ecological niche and grow. Therefore, this observation might be another argument for the universal principle underlying microbial colonization of CF respiratory tracts.

The origin of these species is of particular interest to evaluate their colonization abilities. Most species observed in CF sputum are either widespread organism, that are commonly found in soil or water, or species known to be normal inhabitants of oral cavities and upper respiratory tracts in humans (The Human Microbiome Project Consortium 2012) (Chapter III). Oral cavities potentially act as a reservoir and 'stepping stone' for bacterial immigration into the CF lung (Rogers et al. 2006). Those immigrations are further reported for distinct clones of *P. aeruginosa* which were demonstrated to migrate from the paranasal sinuses, where they evolve and gain growth advantages, into lower airways to which they had become well adapted (Hansen et al. 2012). Likewise, the oral fungal microbiome is known to consist not only of bacteria but also of different yeast species, in particular from the genus *Candida* (Ghannoum et al. 2010). *Saccharomycetes* and other yeasts also make a large proportion of fungi observed in sputum samples (Müller & Seidler 2010). Similar migration routes are suggested for these species (Chapter III). For most other fungi, aerial dispersion is more likely the driver of distribution (Fröhlich-Nowoisky et al. 2009). Considering the higher number of fungal species, in comparison to bacteria, observed in a broad CF cohort and a considerably large number of species only detected in one sample, the colonization abilities of these fungi are assumingly not as adapted to the lung environment as those of bacteria (Chapter III).

In this regard, resilience of microbial communities has informative character about the colonization abilities. Dynamics of communities were analyzed and revealed great differences between fungi and bacteria considering the number of species observed in repeatedly collected sputum samples from one patient (Chapter III). Bacterial richness is rather constant in comparison to fungi and even the bacterial community composition exhibits great stability over time (Chapter II). Therefore, bacterial communities are apparently less influenced by environmental factors, like seasonal variation and its accompanied change in vegetation or the change in life styles of patients. All human body sites and the respiratory tract in particular, are exposed to different species during the year. The vegetation plays an obvious role in this

change but also the different activities of the human hosts in different seasons have to be considered. Certain numbers of bacteria are always present in the sputum, whereas fungi are generally less represented or even not detectable (Chapter III). Therefore, greater variation in the respective richness for individual samples might be explained by the changing exposition in different seasons of the year.

However, whereas bacterial colonization of the lower respiratory tracts in CF patients might be considered to be permanent, persistent fungal colonization assumingly depends on special conditions which only occur in rare cases (Chapter II) (Chapter III). Therefore, the majority of fungi is rather transient than persistent in the CF airways. In contrast, bacterial species assumingly find a broader range of ecological niches. Considering the presence of a thick mucus layer with a resulting oxygen gradient, even anaerobic conditions and respective ecological niches for anaerobic bacteria are logical consequences. Accordingly, compositional analysis even revealed persistently high relative abundances of strict anaerobic species in sputum samples of CF patients, like the *S. millerii* group or *F. nucleatum* (Chapter II). Thus, also infrequently observed bacteria can become dominant members of microbial communities under respective conditions and might influence the health of the host.

5.3 Impact of communities on microbial infections

The exploration of the human microbiome shifted the point of view on infectious research. Not only the introduction of a highly pathogenic organism to the human body causes a disease, likewise important seems to be the response of the microbial community to the presence of the intruder (Casadevall & Pirofski 2000) (Jenkinson & Lamont 2005). To understand the processes accompanying an infection, to initiate the correct treatment and to predict confidently the outcome of a disease, the importance of microbial community profiling is becoming more apparent. Because of the diagnostic challenges, the impact of the microbial community on lower respiratory infections in CF patients might be under-evaluated. Accordingly, recent molecular studies elucidated a growing number of emerging pathogens and a broad spectrum of bacteria associated with CF sputum, including anaerobic species (Tunney et al. 2008) (Bittar et al. 2008) (Rogers et al. 2009) (Guss et al. 2011). A general correlation between the colonization by anaerobic bacteria and *P. aeruginosa* infections was suggested but respective studies revealed ambiguous results (Tunney et al. 2008) (Worlitzsch et al. 2009) (LiPuma 2010). However, in communities with persistently dominant anaerobes, *P. aeruginosa* frequently appeared to be more abundant (Chapter II). In this context, the

observation of increased co-colonization from *P. aeruginosa* and *S. millerii* species in CF is of particular interest (Parkins et al. 2008). Enhanced virulence was suggested for this bacterial consortium. How such mutual influences on the growing behaviour of different species might occur is further demonstrated for *F. nucleatum*, which forms co-aggregates and mixed biofilms (Bolstad et al. 1996). In oral cavities, it enables growth of other pathogens by creating specialized microenvironments and benefits from exchange of metabolites (Diaz et al. 2002). *S. millerii* and *F. nucleatum* were both observed in persistently high relative abundances together with *P. aeruginosa* in CF sputum (Chapter II). The concept of polymicrobial consortia as one pathogenic entity was introduced by Rogers et al. (Rogers et al. 2010) in CF research. Consequentially, the entire microbial community may be regarded as opportunistically pathogenic (Jenkinson & Lamont 2005). In this regard, strongly reduced lung functions may not always be correlated with the abundances of single pathogens and health conditions may be better predicted by the determination of polymicrobial consortia (Chapter II). *P. aeruginosa* as the major pathogen assumingly has a great impact on these consortia. This microorganism is not only known to adapt well to the environment of the CF lung, with the mucoid phenotype being the most studied adaptation, also several mutual interactions with other microorganisms are reported (Hauser et al. 2011). Respective bacterial studies were described previously but besides bacteria also the growth of fungal species is influenced by *P. aeruginosa*. Several studies included the yeast *Candida albicans* which mutually interferes in biosynthetic pathways and growth of *P. aeruginosa* (Cugini et al. 2007) (McAlester et al. 2008) (De Sordi & Mühlschlegel 2009). Furthermore, also the biofilm formation of the filamentous fungi *Aspergillus fumigatus* can be altered by the bacterial CF pathogen (Mowat et al. 2010). All mentioned species were observed in CF sputum, thus it can be assumed that respective colonization abilities of fungi and bacteria are influenced by each other (Chapter III). Consequentially, also fungi have to be included in the polymicrobial approach to study lower respiratory infections. The still unknown pathogenicity of most fungal species in CF illustrates the relatively poor knowledge about these pathogens (Pihet et al. 2009). However, considering the low number of respective reports and the high exposure rate, their pathogenic potential can assumingly considered to be rather low. Further ecological studies might reveal the conditions that promote fungal growth and increase the risk of fungal pulmonary infections.

5.4 Outlook and Conclusion

Elucidation of 72 sputum samples from 56 CF patients revealed that individual respiratory microbial communities are rather unique (Chapter II) (Chapter III). The number of abundant bacteria is comparably small regarding other environmental habitats but community structure and composition vary strongly between CF patients. Little is known about the individual drivers of colonization and respective ecological studies will substantially improve our knowledge about lower respiratory infections. For fungi, high exposure rates together with the large number of observed species is disproportionate to the reports of respective pulmonary infections which suggest a generally low colonization and infection ability for these organisms (Chapter III). Elucidation of the respective conditions to promote fungal growth would enable a more effective prevention of those life-threatening infections. Mutual interactions between bacteria facilitate polymicrobial consortia which in turn might influence substantially the virulence of species. Determination of these consortia appears to be challenging but may reveal and explain new emerging pathogens not considered previously, such as *F. nucleatum* (Chapter II). These observations will further influence the clinical microbiological laboratories. Routine diagnostics detect well the major pathogens, mainly *Proteobacteria* and *S. aureus*, but only allow a limited view in the microbial community (Chapter II). NGS technology will facilitate a more profound knowledge and assumingly broaden the spectrum of microorganisms to be detected in routine diagnostics. The translation of the technology itself into routine diagnostics appears to be challenging due to time und labour consuming preparations or analyses. However, alternatives for conventional based diagnostics are needed and exhaled breath analyses might be an appropriate candidate. Pathogen discrimination *in vitro* together with a first successful implementation in the clinics was promising (Chapter IV). Considering the feasibility of exhaled air analyses based on VOCs in routine diagnostics of CF lung diseases, major challenges have to be coped with. Origins and physiological functions of VOCs have to be elucidated to define their discriminative potential. Breath sampling methods as well as extraction of exhaled VOCs have to be improved and standards developed which are adapted to clinical needs, like the ease of operation.

Overall, compositional microbial studies including bacteria and fungi will further increase the knowledge about the pathogenicity of species and will enable a better prediction on the outcome of infections. In particular, the identification of pathogenic polymicrobial consortia may be a crucial step in the development of advanced diagnostics for respiratory infections.

Critical factors that promote growth of the individual pathogens will be elucidated with a more profound knowledge of the human airway microbial ecology and together with innovative diagnostic approaches, such as exhaled breath analysis, it will further improve the life expectancy and quality of CF patients.

5.5 References

Bittar F, Richet H, Dubus J-C, Reynaud-Gaubert M, Stremler N, Sarles J, Raoult D & Rolain J-M (2008) Molecular Detection of Multiple Emerging Pathogens in Sputa from Cystic Fibrosis Patients. *PLoS ONE* 3: e2908.

Bittar F & Rolain J-M (2010) Detection and accurate identification of new or emerging bacteria in cystic fibrosis patients. *Clin. Microbiol. Infect.* 16: 809–820.

Bolstad AI, Jensen HB & Bakken V (1996) Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. *Clin Microbiol Rev* 9: 55–71.

Casadevall A & Pirofski L-A (2000) Host-Pathogen Interactions: Basic Concepts of Microbial Commensalism, Colonization, Infection, and Disease. *Infect. Immun.* 68: 6511–6518.

Cugini C, Calfee MW, Farrow JM, Morales DK, Pesci EC & Hogan DA (2007) Farnesol, a common sesquiterpene, inhibits PQS production in Pseudomonas aeruginosa. *Mol. Microbiol* 65: 896–906.

Diaz PI, Zilm PS & Rogers AH (2002) Fusobacterium nucleatum supports the growth of Porphyromonas gingivalis in oxygenated and carbon-dioxide-depleted environments. *Microbiology* 148: 467–472.

Fröhlich-Nowoisky J, Pickersgill DA, Després VR & Pöschl U (2009) High diversity of fungi in air particulate matter. *Proc Natl Acad Sci U S A* 106: 12814–12819.

Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature* 459: 193–199.

Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A & Gillevet PM (2010) Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathog* 6: e1000713.

Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL & Singh PK (2012) Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 109: 13769–13774.

Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S & Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 5: 20–29.

Hansen SK et al. (2012) Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *ISME J* 6: 31–45.

Hauser AR, Jain M, Bar-Meir M & McColley SA (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clin Microbiol Rev* 24: 29–70.

Jenkinson HF & Lamont RJ (2005) Oral microbial communities in sickness and in health. *Trends Microbiol.* 13: 589–595.

LiPuma JJ (2010) The Changing Microbial Epidemiology in Cystic Fibrosis. *Clin. Microbiol. Rev.* 23: 299–323.

McAlester G, O'Gara F & Morrissey JP (2008) Signal-mediated interactions between Pseudomonas aeruginosa and Candida albicans. *J. Med. Microbiol* 57: 563–569.

Mowat E, Rajendran R, Williams C, McCulloch E, Jones B, Lang S & Ramage G (2010) Pseudomonas aeruginosa and their small diffusible extracellular molecules inhibit Aspergillus fumigatus biofilm formation. *FEMS Microbiology Letters* 313: 96–102.

Müller F-MC & Seidler M (2010) Characteristics of pathogenic fungi and antifungal therapy in cystic fibrosis. *Expert Rev Anti Infect Ther* 8: 957–964.

Paredi P, Kharitonov SA, Leak D, Shah PL, Cramer D, Hodson ME & Barnes PJ (2000) Exhaled ethane is elevated in cystic fibrosis and correlates with carbon monoxide levels and airway obstruction. *Am. J. Respir. Crit. Care Med.* 161: 1247–1251.

Paredi P, Shah PL, Montuschi P, Sullivan P, Hodson ME, Kharitonov SA & Barnes PJ (1999) Increased carbon monoxide in exhaled air of patients with cystic fibrosis. *Thorax* 54: 917–920.

Parkins MD, Sibley CD, Surette MG & Rabin HR (2008) The Streptococcus milleri group--an unrecognized cause of disease in cystic fibrosis: a case series and literature review. *Pediatr. Pulmonol.* 43: 490–497.

Pedrós-Alió C (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* 4: 449–466.

Pihet M et al. (2009) Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis – a review. *Medical Mycology* 47: 387–397.

Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P, Tunlid A & Hagström Å (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology* 16: 867–880.

Raoult D, Fournier PE & Drancourt M (2004) What does the future hold for clinical microbiology? *Nat. Rev. Microbiol.* 2: 151–159.

Robroeks CMHHT et al. (2010) Metabolomics of Volatile Organic Compounds in Cystic Fibrosis Patients and Controls. *Pediatric Research* 68: 75–80.

Rogers GB, Carroll MP & Bruce KD (2009) Studying bacterial infections through culture-independent approaches. *J Med Microbiol* 58: 1401–1418.

Rogers GB, Carroll MP, Hoffman LR, Walker AW, Fine DA & Bruce KD (2010) Comparing the microbiota of the cystic fibrosis lung and human gut. *Gut Microbes* 1: 85–93.

Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ & Bruce KD (2006) Use of 16S rRNA Gene Profiling by Terminal Restriction Fragment Length Polymorphism Analysis To Compare Bacterial Communities in Sputum and Mouthwash Samples from Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 44: 2601–2604.

Rudkjøbing VB, Thomsen TR, Alhede M, Kragh KN, Nielsen PH, Johansen UR, Givskov M, Høiby N & Bjarnsholt T (2011) True Microbiota Involved in Chronic Lung Infection of Cystic Fibrosis Patients Found by Culturing and 16S rRNA Gene Analysis ▽. *J Clin Microbiol* 49: 4352–4355.

De Sordi L & Mühlschlegel FA (2009) Quorum sensing and fungal–bacterial interactions in Candida albicans: a communicative network regulating microbial coexistence and virulence. *FEMS Yeast Research* 9: 990–999.

Stressmann FA et al. (2012) Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* 67: 867–873.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207–214.

Tunney MM et al. (2008) Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* 177: 995–1001.

Willner D, Haynes MR, Furlan M, Schmieder R, Lim YW, Rainey PB, Rohwer F & Conrad D (2012) Spatial distribution of microbial communities in the cystic fibrosis lung. *ISME J* 6: 471–474.

Worlitzsch D, Rintelen C, Böhm K, Wollschläger B, Merkel N, Borneff-Lipp M & Döring G (2009) Antibiotic-resistant obligate anaerobes during exacerbations of cystic fibrosis patients. *Clin. Microbiol. Infect.* 15: 454–460.

Zhao J et al. (2012) Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci. U.S.A.* 109: 5809–5814.

Appendix

1. Post-print of an article published in

Rolf Kramer & Wolf-Rainer Abraham,
Volatile sesquiterpenes from fungi: What are they good for?
(2012) Phytochemistry Reviews, 11, pp. 15-37.
DOI: 10.1007/s11101-011-9216-2

2. Curriculum Vitae

Rolf Kramer
Sack 13, 38100 Braunschweig
Mobil (+49) 0151 232 632 62
email kramer-rolf@gmx.de

Geburtsdatum 14.12.1981
Geburtsort Papenburg, Niedersachsen

CURRICULUM VITAE

Mai 2009 - 2013	**Anfertigung der Promotion in der Mikrobiologie ,** Helmholtz Zentrum für Infektionsforschung (HZI), Abteilung für Angewandte Mikrobiologie und Vakzinologie; Braunschweig • *Doktorarbeit*: "Elucidation of respiratory microbial communities and the feasibility of breath diagnostics based on microbial volatile organic compounds (VOCs) in adult cystic fibrosis patients"
April 2008 - 2009	**Auslandsaufenthalt,** Australien, Neuseeland, Südostasien
Jan. 2008 - März 2008	**wissenschaftliche Hilfskraft,** Chemotherapeutisches Forschungsinstitut (Georg-Speyer-Haus), Abteilung für Stammzellbiologie und Stammzelltherapie; Frankfurt am Main • *Projekt:* "Analyse proviraler Integrationsloci in der Gentherapie"
April 2002 - Dez. 2007	**Diplom im Fach Biologie,** Schwerpunkte in Genetik, Immunologie und Zoologie Johannes Gutenberg Universität Mainz • *Diplomarbeit:* "Analyse proviraler DNA-Methylierung in einer Gentherapie-Studie zur Behandlung von X-CGD" angefertigt im „Chemotherapeutisches Forschungsinstitut (Georg-Speyer-Haus) in Frankfurt am Main", Abteilung für Stammzellbiologie und Stammzelltherapie • *Schwerpunkte:* Epigenetische Analysen, Isolierung mononukleärer Zellen aus humanem Blut, *ex vivo* Zellkulturen hämatopoetischer Zellen
2001 - März 2002	**ungelernte Hilfskraft im Handwerk,** zur Finanzierung des Studiums
1988 - 2001	**Allgemeine Hochschulreife,** Kreisgymnasium Haren (Ems); Niedersachsen

ZUSATZQUALIFIKATIONEN

2010/2011 *einjährig*	**Doktorandensprecher** des Helmholtz Zentrums für Infektionsforschung
2009-2012 *ganzjährig*	**Helmholtz International Graduate School for Infection Research**
2011/2012 *je Modul 1-2 Tage*	**Management Curriculum** *Themen:* Teammanagement, Mitarbeiterführung, Projektmanagement, Grundlagen der Betriebswirtschaftslehre, Moderation in Besprechungen, Verhandlungstraining
06.2012 *5-Tages Workshop*	**Statistical course: Multivariate Analysis** using Primer v6 *Themen:* Statistische Analysen und deren Interpretation, Multi-Dimensional Scaling (MDS), Principal Components Analysis (PCA); in Plymouth, UK
11.2011 *2-Tages Workshop*	**Research Management Training Workshop (ReMaT)** given by Brussels Office of the Helmholtz Association of German Research Centres *Themen:* Management interdisziplinärer Projekte, Beantragung wissenschaftlicher Fördergelder in Europa, Patentrecht und Innovation

BESONDERE KENNTNISSE

Sprachen

- Englisch, fließend in Wort und Schrift
- Französisch, Grundkenntnisse

EDV

- MS Office; Word, Excel, Power Point
- verschiedene wissenschaftliche Statistikprogramme

WISSENSCHAFTLICHE PUBLIKATIONEN

Kramer, R.; Abraham W.R. (2012) Volatile sesquiterpenes from fungi: What are they good for? (Review article*, **Phytochemistry Reviews. 11, 15-17**)*